(51) International Patent Classification 5 : C07K 15/04, A61K 39/29 G01N 33/569, C07K 7/10, 7/08 C07K 7/06	A2	(11) International Publication Number: WO 93/00365
		(43) International Publication Date: 7 January 1993 (07.01.93)

(21) International Application Number: PCT/US92/05388

(22) International Filing Date: 24 June 1992 (24.06.92)

(30) Priority data:
722,489   24 June 1991 (24.06.91)   US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors: CHIEN, David, Y. ; 1121 Douglas Court, Alamo, CA 94507 (US). RUTTER, William ; 80 Everson Street, San Francisco, CA 94131 (US).

(74) Agents: BLACKBURN, Robert, P. et al.; Chiron Corporation, Intellectual Property Department, 4560 Horton Street, Emeryville, CA 94608 (US).

(81) Designated States: AU, BG, CA, FI, HU, JP, KR, NO, PL, RO, RU, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HEPATITIS C VIRUS (HCV) POLYPEPTIDES

(57) Abstract

HCV epitopes useful as immunogenic reagents, and corresponding antibodies and methods of use are disclosed.

5   1

HEPATITIS C VIRUS (HCV) POLYPEPTIDES

10   Technical Field

            The invention relates to materials and methodologies for managing
the spread of hepatitis C virus (HCV) infection. More specifically, it relates to
polypeptides useful as immunological reagents in the detection, prevention and
treatment of HCV infections.

15

Background

            HCV was first identified and characterized as a cause of non-A,
non-B hepatitis (NANBH) by Houghton et al. This led to the disclosure of a
number of general and specific polypeptides useful as immunological reagents.

20   *See, e.g.,* Houghton et al., EPO Pub. No. 318,216; Houghton et al., EPO Pub.
No. 388,232; Choo et al., Science (1989) 244:359-362; Kuo et al., Science (1989)
244:362-364; Houghton et al., Hepatology (1991) 14:381-388. These
publications provide the art with an extensive background on HCV generally, as
well as the manufacture and uses of HCV polypeptide immunological reagents.

25   For brevity, therefore, the disclosure of these publications in particular are
incorporated herein by reference.

            Others have readily applied and extended the work of Houghton et
al. *See, e.g.,* Highfield et al., UK Pat. App. 2,239,245 (The Wellcome
Foundation Ltd.); Wang, EPO Pub. No. 442,394 (United Biomedical Inc.); Leung

30   et al., EPO Pub. No. 445,423 (Abbott Laboratories); Habits et al., EPO Pub. No.
451,891 (Akzo N.V.); Reyes et al., PCT Pub. No. WO 91/15516 (Genelabs Inc.);

2

Maki et al., EPO Pub. No. 468,657 (Tonen Corp.); and Kamada et al., EPO Pub. No. 469,348 (Shionogi Seiyaku K.K.).

Sensitive, specific methods for screening and identifying carriers of HCV and HCV-contaminated blood or blood products are an important advance in

5   medicine. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and HCV has accounted for up to 90% of these cases. The major problem in this disease is the frequent progression to chronic liver damage (25-55%).   Patient care as well as the prevention of transmission of HCV by blood and blood products or by close personal contact require reliable diagnostic

10   and prognostic tools, such as HCV polypeptides, to detect antibodies related to HCV.   Such polypeptides are also useful as vaccines and immunotherapeutic therapeutic agents for the prevention and/or treatment of the disease.

Since HCV is a relatively new agent, a continuing need exists to define additional immunological reagents that will allow further study of the

15   clinical course of disease and the epidemiology of HCV in the population.

Disclosure of the Invention

The invention pertains to the characterization of new HCV epitopes. The characterization of these epitopes permits the manufacture of polypeptide

20   products which reacted immunologically with antibodies to HCV and/or generate anti-HCV antibody production in vivo. These polypeptide products are useful as standards or reagents in diagnostic tests and/or as components of vaccines. Antibodies, including for example both polyclonal and monoclonal, directed against HCV epitopes contained within these polypeptide sequences are also useful

25   reagents, for example, in diagnostic tests, as therapeutic agents, for screening of antiviral agents, and for the isolation of isolation/purification of HCV polypeptides or particles.

In its broadest sense, the present invention is directed to polypeptides containing the newly characterized HCV epitopes disclosed herein,

30   methods of manufacturing such polypeptides (e.g., recombinant and synthetic methods), methods of using such polypeptides (e.g., diagnostic, vaccine, and therapeutic), and articles of manufacture, compositions or formulations adapted to

such uses (e.g., polypeptides fixed to an immunoassay or other support, oral or injectable pharmaceutical compositions). Similarly, antibodies (polyclonal, monoclonal, or equivalents such as binding fragments, single-chain antigen-binding proteins, etc.) to the HCV epitopes disclosed herein are also included within the

5 scope of the present invention, as well as methods of making such antibodies, methods of using such antibodies (e.g., diagnostic, vaccine, and therapeutic), and articles of manufacture, compositions or formulations adapted to such uses (e.g., antibodies fixed to an immunoassay or other support, oral or injectable pharmaceutical compositions).

10 Other aspects of the invention pertain to kits for analyzing samples for the presence of an HCV antigen comprising the above antibodies in a suitable container. Still other aspects of the invention pertain to kits for analyzing samples for the presence of an antibodies directed against an HCV antigen comprising a polypeptide as described above in a suitable container.

15 Still other aspects of the invention are: a method for producing a polypeptide containing the newly disclosed HCV epitopes comprising incubating host cells transformed with an expression vector containing a sequence encoding a polypeptide containing the HCV epitope under conditions which allow expression of said polypeptide; and a polypeptide containing such an HCV epitope produced

20 by this method.

Immunoassays are also included in the invention. These include an immunoassay for detecting an HCV antigen comprising incubating a sample suspected of containing an HCV antigen with an antibody as described above under conditions which allow the formation of an antigen-antibody complex; and

25 detecting an antigen-antibody complex containing the antibody. An immunoassay for detecting anti-HCV antibodies comprising incubating a sample suspected of containing anti-HCV antibodies with a polypeptide as described above, under conditions which allow the formation of an antibody-antigen complex; and detecting the antibody-antigen complex containing the polypeptide.

30 Also included in the invention are vaccines for treatment of HCV infection comprising an immunogenic peptide containing an HCV epitope described herein.

4

Yet another aspect of the invention is a method for producing antibodies to HCV comprising administering to an individual an isolated immunogenic polypeptide containing an HCV epitope described herein in an amount sufficient to produce an immune response.

5   The above aspects of the present invention are accomplished by the discovery of HCV epitopes of the formula
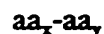
$$aa_x\text{-}aa_y$$

wherein aa denotes an amino acid;

x and y are integers such that $y\text{-}x \geq 6$;

10         $aa_x\text{-}aa_y$ indicates a portion of the amino acid sequence of Figure 1; and

x is selected from the group consisting of 23-34, 36, 66-79, 81-94, 96-98, 101-103, 186-189, 191, 206, 223, 232, 256, 286, 297-299, 321, 347, 357, 413, 414, 432, 465-471, 480-484, 501, 502, 521, 540-549, 579, 594-599, 601-
15   613, 641, 662-665, 685, 705, 706, 729, 782-789, 801, 851-855, 893, 916, 928, 946, 952-954, 1026, 1072, 1109, 1112-1117, 1218, 1240, 1280-1285, 1322, 1338, 1371, 1384, 1410, 1411, 1454, 1492, 1493, 1532-1535, 1560, 1561, 1566-1568, 1571-1577, 1601-1607, 1615-1620, 1655, 1695, 1710-1712, 1728, 1729, 1758-1762, 1781, 1808, 1821, 1851, 1880, 1908-1913, 1925, 1940-1948, 1951, 1966-
20   1969, 1999, 2001-2004, 2006-2014, 2024, 2048-2053, 2055-2057, 2071, 2088-2093, 2108, 2122-2148, 2165, 2187, 2226-2232, 2244-2249, 2267, 2281-2286, 2288, 2289, 2325-2327, 2346, 2347, 2349, 2382, 2401, 2417-2422, 2439-2444, 2446-2456, 2469, 2471-2476, 2495, 2533, 2534, 2573-2578, 2602-2604, 2606-2612, 2632-2638, 2660, 2676-2679, 2688-2693, 2707, 2721, 2757-2762, 2779,
25   2794, 2795, 2797-2799, 2801, 2802, 2817-2843, 2863-2867, 2878-2884, 2886-2895.

5

The above objects are also achieved using HCV epitopes of the
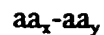
formula

$$aa_x\text{-}aa_y$$

wherein aa denotes an amino acid;

5   x and y are integers such that $y\text{-}x \geq 6$;

$aa_x\text{-}aa_y$ indicates a portion of the amino acid sequence of Figure 1;

and   x is selected from the group consisting of 35 (where y is less than

45), 80 (where y is less than 90), 95 (where y is less than 110), 99 (where y is

less than 120), 100 (where y is less than 150), 190 (where y is less than 210), 500

10   (where y is less than 550), 600 (where y is less than 625), 1260 (where y is less

than 1280), 1569 (where y is less than 1931), 1570 (where y is less than 1590),

1694 (where y is less than 1735), 1949 (where y is less than 2124), 1950 (where y

is less than 1985), 2000 (where y is less than 2050), 2005 (where y is less than

2025), 2054 (where y is less than 2223), 2250 (where y is less than 2330), 2287

15   (where y is less than 2385), 2290 (where y is less than 2310), 2345 (where y is

less than 2375), 2348 (where y is less than 2464), 2445 (where y is less than

2475), 2470 (where y is less than 2490), 2605 (where y is less than 2620), 2780

(where y is less than 2830), 2796 (where y is less than 2886), 2800 (where y is

less than 2850), and 2885 (where y is less than 2905).

20   In either of the above formula, x-y can less than or equal to 10, 20,

30, 40 or 50 in some embodiments of the invention.

Brief Description of the Drawings

Fig. 1 shows the polyprotein of the HCV prototype isolate HCV1.

25   Fig. 2 shows a composite cDNA sequence for HCV1.

Fig. 3 shows the nucleotide consensus sequence of human isolate

23, variant sequences are shown below the sequence line. The amino acids

encoded in the consensus sequence are also shown.

Fig. 4 shows the nucleotide consensus sequence of human isolate

30   27, variant sequences are shown below the sequence line. The amino acids

encoded in the consensus sequence are also shown.

6

Fig. 5 shows the aligned nucleotide sequences of human isolates 23 and 27 and of HCV1. Homologous sequences are indicated by the symbol (*). Non-homologous sequences are in small letters.

Fig. 6 shows the aligned amino acid sequences of human isolates 23
5   and 27 and of HCV1. Homologous sequences are indicated by the symbol (*). Non-homologous sequences are in small letters.

Fig. 7 shows a comparison of the composite aligned nucleotide sequences of isolates Thorn, EC1, HCT #18, and HCV1.

Fig. 8 shows a comparison of the nucleotide sequences of EC10 and
10   a composite of the HCV1 sequence; the EC10 sequence is on the line above the dots, and the HCV1 sequence is on the line below the dots.

Fig. 9 shows a comparison of the amino acid sequences 117-308 (relative to HCV1) encoded in the "EnvL" regions of the consensus sequences of human isolates HCT #18, JH23, JH 27, Thorne, EC1, and of HCV1.

15   Fig. 10 shows a comparison of the amino acid sequences 330-360 (relative to HCV1) encoded in the "EnvR" regions of the consensus sequences of human isolates HCT #18, JH23, JH 27, Thorne, EC1, and of HCV1.

Modes for Carrying Out the Invention
20

Complete citations to publications referred to herein can be found in the "Background" or "Bibliography" sections.

I. Definitions

"Hepatitis C virus" or "HCV" refers to the art-recognized viral
25   species of which pathogenic strains cause NANBH, and attenuated strains or defective interfering particles derived therefrom. See generally, publications cited in the section entitled "Background." The HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation, $i.e.$, reportedly on the order of 10^{-3} to 10^{-4} per incorporated nucleotide
30   (Fields & Knipe (1986)). Therefore, since heterogeneity and fluidity of genotype are inherent in RNA viruses, there are multiple strains/isolates, which may be virulent or avirulent, within the HCV species. The propagation, identification,

SUBSTITUTE SHEET

detection, and isolation of the various HCV strains or isolates has been well documented in the literature. Moreover, the disclosure herein allows the preparation of diagnostics and vaccines for the various strains/isolates, as well as compositions and methods that have utility in screening procedures for anti-viral

5   agents for pharmacologic use, such as agents that inhibit replication of HCV.

Information on several different strains/isolates of HCV is disclosed herein, particularly strain or isolate CDC/HCV1 (also called HCV1). Information from one strain or isolate, such as a partial genomic or amino acid sequence, is sufficient to allow those skilled in the art using standard techniques to isolate new

10 strains/isolates and to identify whether such new strains/isolates are HCV. For example, several different strains/isolates are described below. These strains, which were obtained from a number of human sera (and from different geographical areas), were isolated utilizing the information from the genomic sequence of HCV1.

15   The information provided herein is indicative that HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of Flavivirus particles are known, and are discussed in Brinton (1986). Generally, with respect to morphology, Flaviviruses contain a central

20 nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections that are about 5-10 nm long with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever virus. They

25 possess positive-stranded RNA genomes (~11,000 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

The genomic structure and the nucleotide sequence of HCV genomic RNA has been deduced. The genome appears to be single-stranded RNA con-

30 taining ~ 10,000 nucleotides. The genome is positive-stranded, and possesses a continuous, translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amin acids. In the ORF, the structural protein(s) appear to be

SUBSTITUTE SHEET

8

encoded in approximately the first quarter of the N-terminus region, with the
majority of the polyprotein responsible for non-structural proteins. When
compared with all known viral sequences, small but significant co-linear
homologies are observed with the non-structural proteins of the flavivirus family,

5   and with the pestiviruses (which are now also considered to be part of the
Flavivirus family).

Based upon the putative amino acids encoded in the nucleotide
sequence of HCV1 and other evidence, possible protein domains of the encoded
HCV polyprotein, as well as the approximate boundaries, are the following:

10

Putative Domain	Approximate Boundary (amino acid nos.)
C (nucleocapsid protein)	1-191
E_1 (virion envelope protein)	192-383
E_2/NS1 (envelope?)	384-800
NS2 (unknown function)	800-1050
NS3 (protease?)	1050-1650
NS4 (unknown function)	1651-2100
NS5 (polymerase)	2100-3011(end)

15

20

25

These domains are, however, tentative. For example, the E1-NS2 border is probably in the 750-810 region, and NS3-NS4 border is about 1640-1650. There is also evidence that the 191 aa version of C is a precursor that is further

5   processed (e.g., to about 170 aa in length), and that the NS2, NS4 and NS5 proteins are each further processed into two mature proteins.

Different strains, isolates or subtypes of HCV are expected to contain variations at the amino acid and nucleic acids compared with HCV1. Many isolates are expected to show much (*i.e.*, more than about 40%) homology

10   in the total amino acid sequence compared with HCV1. However, it may also be found that there are other less homologous HCV isolates. These would be defined as HCV according to various criteria such as, for example, an ORF of approximately 9,000 nucleotides to approximately 12,000 nucleotides, encoding a polyprotein similar in size to that of HCV1, an encoded polyprotein of similar

15   hydrophobic and/or antigenic character to that of HCV1, and the presence of co-linear peptide sequences that are conserved with HCV1. In addition, the genome would be a positive-stranded RNA.

HCV encodes at least one epitope which is immunologically identifiable with an epitope in the HCV1 polyprotein. The epitope is unique to

20   HCV when compared to previously known Flaviviruses. The uniqueness of the epitope may be determined by its immunological reactivity with anti-HCV antibodies and lack of immunological reactivity with antibodies to known Flavivirus species. Methods for determining immunological reactivity are known in the art, for example, by radioimmunoassay, by ELISA assay, by hemagglutination, and

25   several examples of suitable techniques for assays are provided herein. Alternatively, a comparison of the sequence of the HCV epitope to previously known sequences of members of the Flavivirus family can be used to evaluate "uniqueness."

In addition to the above, the following parameters of nucleic acid

30   homology and amino acid homology are applicable, either alone or in combination, in identifying a strain/isolate as HCV. Since HCV strains and isolates are evolutionarily related, it is expected that the overall homology of the genomes at the nucleotide level may be about 10% or greater, probably will be

about 40% or greater, probably about 60% or greater, and even more probably about 80% or greater; and in addition that there will be corresponding contiguous sequences of at least about 13 nucleotides. It should be noted that variable and hypervariable regions within the HCV genome; therefore, the homology in these

5 regions is expected to be significantly less than that in the overall genome. The correspondence between the putative HCV strain genomic sequence and, for example, the CDC/HCV1 cDNA sequence can be determined by techniques known in the art. For example, they can be determined by a direct comparison of the sequence information of the polynucleotide from the putative HCV, and the HCV

10 cDNA sequence(s) described herein. For example, also, they can be determined by hybridization of the polynucleotides under conditions which form stable duplexes between homologous regions (for example, those which would be used prior to S_1 digestion), followed by digestion with single stranded specific nuclease(s), followed by size determination of the digested fragments.

15   Because of the evolutionary relationship of the strains or isolates of HCV, putative HCV strains or isolates are identifiable by their homology at the polypeptide level. Generally, HCV strains or isolates are expected to be at least 10% homologous, more than about 40% homologous, probably more than about 70% homologous, and even more probably more than about 80% homologous, and

20 some may even be more than about 90% homologous at the polypeptide level. The techniques for determining amino acid sequence homology are known in the art. For example, the amino acid sequence may be determined directly and compared to the sequences provided herein. Alternatively the nucleotide sequence of the genomic material of the putative HCV may be determined (usually via a

25 cDNA intermediate), the amino acid sequence encoded therein can be determined, and the corresponding regions compared.

        As used herein, a polynucleotide "derived from" a designated sequence refers to a polynucleotide sequence which is comprised of a sequence of approximately at least about 6 nucleotides, preferably at least about 8 nucleotides,

30 more preferably at least about 10-12 nucleotides, and even more preferably at least about 15-20 nucleotides corresponding to a region of the designated nucleotide sequence. "Corresponding" means homologous to or complementary to the designated sequence. Preferably, the sequence of the region from which the

polynucleotide is derived is homologous to or complementary to a sequence which is unique to an HCV genome. Whether or not a sequence is unique to the HCV genome can be determined by techniques known to those of skill in the art. For example, the sequence can be compared to sequences in databanks (as of the

5   priority date), e.g., Genebank, to determine whether it is present in the uninfected host or other organisms. The sequence can also be compared to the known (as of the priority date) sequences of other viral agents, including those which are known to induce hepatitis, e.g., HAV, HBV, and HDV, and to members of the Flaviviridae. The correspondence or non-correspondence of the derived sequence

10   to other sequences can also be determined by hybridization under the appropriate stringency conditions. Hybridization techniques for determining the complementarity of nucleic acid sequences are known in the art. See, for example, Maniatis et al. (1982). In addition, mismatches of duplex poly-nucleotides formed by hybridization can be determined by known techniques,

15   including for example, digestion with a nuclease such as S1 that specifically digests single-stranded areas in duplex polynucleotides. Regions from which typical DNA sequences may be "derived" include but are not limited to, for example, regions encoding specific epitopes, as well as non-transcribed and/or non-translated regions.

20   The derived polynucleotide is not necessarily physically derived from the nucleotide sequence shown, but may be generated in any manner, including for example, chemical synthesis or DNA replication or reverse transcription or transcription. In addition, combinations of regions corresponding to that of the designated sequence may be modified in ways known in the art to be

25   consistent with an intended use.

          Similarly, a polypeptide or amino acid sequence "derived from" a designated amino acid or nucleic acid sequence refers to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least 3-5 amino acids, and

30   more preferably at least 8-10 amino acids, and even more preferably at least 11-15 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence. This terminology also includes a polypeptide expressed from a designated nucleic acid sequence.

A recombinant or derived polypeptide is not necessarily translated
from a designated nucleic acid sequence; it may be generated in any manner,
including for example, chemical synthesis, or expression of a recombinant
expression system, or isolation from HCV, including mutated HCV. A recombin-
ant or derived polypeptide may include one or more analogs of amino acids or
unnatural amino acids in its sequence. Methods of inserting analogs of amino
acids into a sequence are known in the art. It also may include one or more
labels, which are known to those of skill in the art.

The term "recombinant polynucleotide" as used herein intends a
polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by
virtue of its origin or manipulation: (1) is not associated with all or a portion of a
polynucleotide with which it is associated in nature, (2) is linked to a
polynucleotide other than that to which it is linked in nature, or (3) does not occur
in nature.

The term "polynucleotide" as used herein refers to a polymeric form
of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This
term refers only to the primary structure of the molecule. Thus, this term
includes double- and single-stranded DNA and RNA. It also includes known types
of modifications, for example, labels which are known in the art, methylation,
"caps", substitution of one or more of the naturally occurring nucleotides with an
analog, internucleotide modifications such as, for example, those with uncharged
linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates,
carbamates, *etc.*) and with charged linkages (*e.g.*, phosphorothioates, phosphoro-
dithioates, *etc.*), those containing pendant moieties, such as, for example proteins
(including for *e.g.*, nucleases, toxins, antibodies, signal peptides, poly-L-lysine,
etc.), those with intercalators (*e.g.*, acridine, psoralen, *etc.*), those containing
chelators (*e.g.*, metals, radioactive metals, boron, oxidative metals, *etc.*), those
containing alkylators, those with modified linkages (*e.g.*, alpha anomeric nucleic
acids, *etc.*), as well as unmodified forms of the polynucleotide.

A "purified" polypeptide refers to the polypeptide being in a state
that is substantially free of other polypeptides, *i.e.*, in a composition that contains
a minimum of about 50% by weight (desired polypeptide/total polypeptide in
composition), preferably a minimum of about 70%, and even more preferably a

minimum of about 90% of the desired polypeptide, without regard to non-proteinaceous materials in the composition. Techniques for purifying viral poly-peptides are known in the art. Purified antibodies are similarly defined.

5   "Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures", and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in

10 morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

A "replicon" is any genetic element, *e.g.*, a plasmid, a chromosome, a virus, a cosmid, *etc.* that behaves as an autonomous unit of polynucleotide replication within a cell; *i.e.*, capable of replication under its own

15 control.

A "vector" is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

"Control sequence" refers to polynucleotide sequences which are

20 necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and terminators; in eukaryotes, generally, such control sequences include promoters, terminators and, in some instances, enhancers. The term "control

25 sequences" is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended

30 manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

An "open reading frame" (ORF) is a region of a polynucleotide sequence which encodes a polypeptide; this region may represent a portion of a coding sequence or a total coding sequence.

A "coding sequence" is a polynucleotide sequence which is

5   transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, and recombinant polynucleotide sequences.

10   "Immunologically identifiable with/as" refers to the presence of epitope(s) and polypeptides(s) which are also present in the designated polypeptide(s), usually HCV proteins. Immunological identity may be determined by antibody binding and/or competition in binding; these techniques are known to those of average skill in the art.

15   As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise 3 or more amino acids that defines the binding site of an antibody. Generally an epitope consists of at least 5 amino acids, and sometimes consists of at least 8 amino acids. Methods of epitope mapping are known in the art.

20   A polypeptide is "immunologically reactive" with an antibody when it binds to an antibody due to antibody recognition of a specific epitope contained within the polypeptide. Immunological reactivity may be determined by antibody binding, more particularly by the kinetics of antibody binding, and/or by competition in binding using as competitor(s) a known polypeptide(s) containing an epitope

25   against which the antibody is directed. The techniques for determining whether a polypeptide is immunologically reactive with an antibody are known in the art.

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one antibody combining site. An "antibody combining site" or "binding domain" is formed from the folding of vari-

30   able domains of an antibody molecule(s) to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows an immunological reaction with the antigen. An antibody combining site may be formed from a heavy and/or a

light chain domain (V_H and V_L, respectively), which form hypervariable loops which contribute to antigen binding. The term "antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, altered antibodies, univalent antibodies, the Fab proteins, and single domain antibodies.

5   As used herein, a "single domain antibody" (dAb) is an antibody which is comprised of an VH domain, which reacts immunologically with a designated antigen. A dAb does not contain a V_L domain, but may contain other antigen binding domains known to exist in antibodies, for example, the kappa and lambda domains. Methods for preparing dAbs are known in the art. See, for

10   example, Ward *et al.* (1989).

Antibodies may also be comprised of V_H and V_L domains, as well as other known antigen binding domains. Examples of these types of antibodies and methods for their preparation are known in the art (see, *e.g.*, U.S. Patent No. 4,816,467, which is incorporated herein by reference), and include the following.

15   For example, "vertebrate antibodies" refers to antibodies which are tetramers or aggregates thereof, comprising light and heavy chains which are usually aggregated in a "Y" configuration and which may or may not have covalent linkages between the chains. In vertebrate antibodies, the amino acid sequences of all the chains of a particular antibody are homologous with the chains found in one

20   antibody produced by the lymphocyte which produces that antibody *in situ*, or *in vitro* (for example, in hybridomas). Vertebrate antibodies typically include native antibodies, for example, purified polyclonal antibodies and monoclonal antibodies. Examples of the methods for the preparation of these antibodies are described *infra*.

25   "Hybrid antibodies" are antibodies wherein one pair of heavy and light chains is homologous to those in a first antibody, while the other pair of heavy and light chains is homologous to those in a different second antibody. Typically, each of these two pairs will bind different epitopes, particularly on different antigens. This results in the property of "divalence", *i.e.*, the ability to

30   bind two antigens simultaneously. Such hybrids may also be formed using chimeric chains, as set forth below.

"Chimeric antibodies", are antibodies in which the heavy and/or light chains are fusion proteins. Typically the constant domain of the chains is

from one particular species and/or class, and the variable domains are from a
different species and/or class. Also included is any antibody in which either or
both of the heavy or light chains are composed of combinations of sequences
mimicking the sequences in antibodies of different sources, whether these sources

5   be differing classes, or different species of origin, and whether or not the fusion
point is at the variable/constant boundary. Thus, it is possible to produce
antibodies in which neither the constant nor the variable region mimic known anti-
body sequences. It then becomes possible, for example, to construct antibodies
whose variable region has a higher specific affinity for a particular antigen, or

10   whose constant region can elicit enhanced complement fixation, or to make other
improvements in properties possessed by a particular constant region.

        Another example is "altered antibodies", which refers to antibodies
in which the naturally occurring amino acid sequence in a vertebrate antibody has
been varied. Utilizing recombinant DNA techniques, antibodies can be redesigned

15   to obtain desired characteristics. The possible variations are many, and range
from the changing of one or more amino acids to the complete redesign of a
region, for example, the constant region. Changes in the constant region, in
general, to attain desired cellular process characteristics, e.g., changes in com-
plement fixation, interaction with membranes, and other effector functions.

20   Changes in the variable region may be made to alter antigen binding
characteristics. The antibody may also be engineered to aid the specific delivery
of a molecule or substance to a specific cell or tissue site. The desired alterations
may be made by known techniques in molecular biology, e.g., recombinant
techniques, site directed mutagenesis, etc.

25   Yet another example are "univalent antibodies", which are
aggregates comprised of a heavy chain/light chain dimer bound to the Fc (i.e.,
constant) region of a second heavy chain. This type of antibody escapes antigenic
modulation. See, e.g., Glennie et al. (1982).

        Included also within the definition of antibodies are "Fab" fragments
30   of antibodies. The "Fab" region refers to those portions of the heavy and light
chains which are roughly equivalent, or analogous, to the sequences which
comprise the branch portion of the heavy and light chains, and which have been
shown to exhibit immunological binding to a specified antigen, but which lack the

effector Fc portion. "Fab" includes aggregates of one heavy and one light chain (commonly known as Fab'), as well as tetramers containing the 2H and 2L chains (referred to as F(ab)$_2$), which are capable of selectively reacting with a designated antigen or antigen family. "Fab" antibodies may be divided into subsets analogous
5   to those described above, i.e, "vertebrate Fab", "hybrid Fab", "chimeric Fab", and "altered Fab". Methods of producing "Fab" fragments of antibodies are known within the art and include, for example, proteolysis, and synthesis by recombinant techniques.

Also included in the term "antibodies" are single-chain antigen-
10   binding (SCA) proteins, such as the type described in the article co-authored by Schlom, J. in the June 15, 1992 issue of Cancer Research (as well as articles cited therein).

As used herein, the term "immunogenic polypeptide" is a polypeptide that elicits a cellular and/or humoral immune response, whether alone
15   or linked to a carrier in the presence or absence of an adjuvant.

The term "polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example,
20   glycosylation, acetylation, phosphorylation and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

25   "Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, f-mating or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

30   "Treatment" as used herein refers to prophylaxis and/or therapy.

An "individual", as used herein, refers to vertebrates, particularly members of the mammalian species, and includes but is not limited to animals

(e.g., dogs, cats, cattle, swine, sheep, goat, rabbits, mice, rats, guinea pigs, etc.), and primates, including monkeys, chimps, baboons and humans.

As used herein, the "sense strand" of a nucleic acid contains the sequence that has sequence homology to that of mRNA. The "anti-sense strand" contains a sequence which is complementary to that of the "sense strand".

As used herein, a "positive stranded genome" of a virus is one in which the genome, whether RNA or DNA, is single-stranded and which encodes a viral polypeptide(s). Examples of positive stranded RNA viruses include Togaviridae, Coronaviridae, Retroviridae, Picornaviridae, and Caliciviridae. Included also, are the Flaviviridae, which were formerly classified as Togaviradae. See Fields & Knipe (1986).

As used herein, "antibody containing body component" refers to a component of an individual's body which is a source of the antibodies of interest. Antibody containing body components are known in the art, and include but are not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, white blood cells, and myelomas.

As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respir-atory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, and also samples of *in vitro* cell culture constituents (including but not limited to conditioned medium resulting from the growth of cells in cell culture medium, putatively virally infected cells, recombinant cells, and cell components).

II. Description of the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, "Molecular Cloning; A Laboratory Manual" (1982); "DNA Cloning, Volumes I and II" (D.N Glover ed. 1985); "Oligonucleotide Synthesis" (M.J. Gait ed, 1984); "Nucleic Acid Hybridization" (B.D. Hames & S.J. Higgins eds. 1984);

"Transcription and Translation" (B.D. Hames & S.J. Higgins eds. 1984);
"Animal Cell Culture" (R.I. Freshney ed. 1986); "Immobilized Cells And
Enzymes" (IRL Press, 1986); B. Perbal, "A Practical Guide To Molecular
Cloning" (1984); the series, "Methods in Enzymology" (Academic Press, Inc.);

5   "Gene Transfer Vectors For Mammalian Cells" (J.H. Miller and M.P. Calos eds.
1987, Cold Spring Harbor Laboratory), Meth Enzymol Vol. 154 and Vol. 155
(Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987),
"Immunochemical Methods In Cell And Molecular Biology" (Academic Press,
London); Scopes, (1987) "Protein Purification: Principles and Practice", Second

10   Edition (Springer-Verlag, N.Y.); and "Handbook of Experimental Immunology",
Volumes I-IV (D.M. Weir and C. C. Blackwell eds 1986). All patents, patent
applications, and publications mentioned herein, both *supra* and *infra*, are hereby
incorporated herein by reference.

15   **II.A.   Truncated HCV Polypeptides**

The useful materials and processes of the present invention are
made possible by the identification below of new HCV epitopes. The knowledge
of these epitopes (or antigenic regions) allows for construction of polypeptides
containing truncated HCV sequences which can be used as immunological

20   reagents.

Truncated HCV amino acid sequences encoding at least one viral
epitope are useful immunological reagents. For example, polypeptides comprising
such truncated sequences can be used as reagents in an immunoassay. These
polypeptides also are candidate subunit antigens in compositions for antiserum

25   production or vaccines. While these truncated sequences can be produced by
various known treatments of native viral protein, it is generally preferred to make
synthetic or recombinant polypeptides comprising an HCV sequence. Polypeptides
comprising these truncated HCV sequences can be made up entirely of HCV
sequences (one or more epitopes, either contiguous or noncontiguous), or HCV

30   sequences and heterologous sequences in a fusion protein. Useful heterologous
sequences include sequences that provide for secretion from a recombinant host,
enhance the immunological reactivity of the HCV epitope(s), or facilitate the
coupling of the polypeptide to an immunoassay support or a vaccine carrier. See,

e.g., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783, the disclosures of which are incorporated herein by reference.

5   The size of polypeptides comprising the truncated HCV sequences can vary widely, the minimum size being a sequence of sufficient size to provide an HCV epitope, while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired HCV epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated HCV amino acid sequence will range from about 5 (or 8)

10   to about 100 amino acids in length. More typically, however, the HCV sequence will be a maximum of about 50 (or 40) amino acids in length, and sometimes a maximum of about 20, 25 or 30 amino acids. It is usually desirable to select HCV sequences of at least about 8, 10, 12 or 15 amino acids.

     Examples of truncated HCV amino acid sequences (octamers) that

15   are useful as described herein are set forth below in the examples. It is to be understood that these peptides do not necessarily precisely map one epitope. Non-immunogenic portions of the sequence can be defined using conventional techniques and deleted from the described sequences. Further, additional truncated HCV amino acid sequences that comprise an epitope or are immunogenic

20   can be identified as described herein.

     Polypeptide products containing the truncated HCV amino acid sequences disclosed below can be prepared as discrete peptides or incorporated into a larger polypeptide, and may find use as described herein. In preferred applications, truncated sequences from the E1 and/or E2 domains have applications

25   in vaccine and therapeutic products. While generally any of the domains can have some diagnostic utility, C, NS3, NS4 and NS5 are particularly preferred, with combinations of C epitopes with epitopes from one or more of the NS3, NS4 or NS5 domains being particularly preferred.

30   II.B. Preparation of Polypeptides

     The availability of DNA sequences encoding HCV amino acid sequences permits the construction of expression vectors encoding antigenically active regions of the polypeptide (See, *e.g.*, Fig. 2). These antigenically active

regions may be derived from coat or envelope antigens or from core antigens, or
from antigens which are non-structural including, for example, polynucleotide
binding proteins, polynucleotide polymerase(s), and other viral proteins required
for the replication and/or assembly of the virus particle. Fragments encoding the

5   desired polypeptides are derived, for example, from viral cDNA clones using
conventional restriction digestion or by synthetic methods, and are ligated into
vectors which may, for example, contain portions of fusion sequences such as
β-Galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and
vectors which are useful for the production of polypeptides which contain fusion

10   sequences of SOD are described in European Patent Office Publication number
0196056, published October 1, 1986. Vectors encoding fusion polypeptides of
SOD and HCV polypeptides, *i.e.*, $NANB_{5-1-1}$, $NANB_{81}$, and C100-3, which is
encoded in a composite of HCV cDNAs, are described in Sections IV.B.1,
IV.B.2, and IV.B.4, respectively. Any desired portion of the HCV cDNA

15   containing an open reading frame (or a synthetic version thereof), can be used to
express a recombinant polypeptide, such as a mature or fusion protein.

        Alternatively, a polypeptide contain HCV epitopes can be provided
by chemical synthesis using standard techniques based on the amino acid sequence
of the figures and the examples.

20   The DNA encoding the desired polypeptide, whether in fused or
unfused form, and whether or not containing a signal sequence to permit secretion,
may be ligated into expression vectors suitable for any convenient host. Both
eukaryotic and prokaryotic host systems are presently used in forming recombinant
polypeptides, and host cell lines is given in EPO Pub. Nos. 318,216. The

25   polypeptide is then isolated from lysed cells or from the culture medium and pur-
ified to the extent needed for its intended use. Purification may be by techniques
known in the art, for example, differential extraction, salt fractionation,
chromatography on ion exchange resins, affinity chromatography, centrifugation,
and the like. See, for example, Methods in Enzymology for a variety of methods

30   for purifying proteins. Such polypeptides can be used as diagnostics, or those
which give rise to neutralizing antibodies may be formulated into vaccines. Anti-
bodies raised against these polypeptides can also be used as diagnostics, or for

passive immunotherapy. In addition, as discussed below, antibodies to these polypeptides are useful for example in isolating and identifying HCV particles.

The HCV polypeptides may also be isolated from HCV virions and truncated (if not already). The virions may be grown in HCV infected cells in

5   tissue culture, or in an infected host.

II.C.   Preparation of Antigenic Polypeptides and Conjugation with Carrier

An antigenic region of a polypeptide is generally relatively small—typically 8 to 10 amino acids or less in length. Fragments of as few as 5

10   amino acids may characterize an antigenic region. These segments may correspond to regions of HCV antigen. Accordingly, using the cDNAs of HCV as a basis, DNAs encoding short segments of HCV polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis.

15   In instances wherein the synthesized polypeptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-

20   pyridylthio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue.) These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an

25   amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, Immun Rev (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid,

30   2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2 nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens

employs the rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

5   Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized Sepharose®, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino

10   acid copolymers; and inactive virus particles, see, for example, Section II.D. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

          In addition to full-length viral proteins, polypeptides comprising

15

II.D.   Preparation of Hybrid Particle Immunogens Containing HCV Epitopes

          The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B

20   surface antigen. See, e.g., US 4,722,840. Constructs wherein the HCV epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles

25   constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and HBV.

          Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in *S. cerevisiae* (P. Valenzuela *et al.* (1982)), as well as in, for example, mammalian cells (P. Valenzuela *et al.* (1984)). The formation of

30   such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath *et al.* (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed

in EPO 174,444, published March 19, 1986; hybrids including heterologous viral

sequences for yeast expression are disclosed in EPO 175,261, published March 26,

1966. These constructs may also be expressed in mammalian cells such as

Chinese hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector

5   (Michelle *et al.* (1984)).

In addition, portions of the particle-forming protein coding sequence

may be replaced with codons encoding an HCV epitope. In this replacement,

regions which are not required to mediate the aggregation of the units to form

immunogenic particles in yeast or mammals can be deleted, thus eliminating

10   additional HBV antigenic sites from competition with the HCV epitope.

II.E.   <u>Preparation of Vaccines</u>

Vaccines may be prepared from one or more immunogenic

polypeptides derived from HCV. These polypeptides may be expressed in

15   various host cells (*e.g.*, bacteria, yeast, insect, or mammalian cells), or

alternatively may be isolated from viral preparations or made synthetically.

Single- or multi-valent vaccines against HCV may be comprised of one or more

epitopes from one or more structural proteins, and/or one or more epitopes from

one or more nonstructural proteins. These vaccines may be comprised of, for

20   example, recombinant HCV polypeptides and/or polypeptides isolated from the

virions. In particular, vaccines are contemplated comprising one or more of the

following HCV proteins, or subunit antigens derived therefrom: E1, E2, C, NS2,

NS3, NS4 and NS5. Particularly preferred are vaccines comprising E1 and/or E2,

or subunits thereof.

25   In addition to the above, it is also possible to prepare live vaccines

of attenuated microorganisms which express one or more recombinant HCV

polypeptides. Suitable attenuated microorganisms are known in the art and

include, for example, viruses (*e.g.*, vaccinia virus (see Brown *et al.* (1986)), as

well as bacteria.

30   The preparation of vaccines which contain an immunogenic

polypeptide(s) as active ingredients, is known to one skilled in the art. Typically,

such vaccines are prepared as injectables, either as liquid solutions or suspensions;

solid forms suitable for solution in, or suspension in, liquid prior to injection may

also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose,

5   glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-

10   D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell

15   wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween® 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing an HCV antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

20   The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or

25   triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form

30   of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

           The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed

with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example,

5   sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

II.F.   Dosage and Administration of Vaccines

10   The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 µg to 250 µg of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the

15   degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and may be peculiar to each subject.

             The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary

20   course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the

25   judgment of the practitioner.

             In addition, the vaccine containing the immunogenic HCV antigen(s) may be administered in conjunction with other immunoregulatory agents, for example, immune globulins.

30   II.G.   Preparation of Antibodies Against HCV Epitopes

             The immunogenic polypeptides described herein are used to produce antibodies, including polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized

with an immunogenic polypeptide bearing an HCV epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an HCV epitope contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity

5   chromatography. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987).Alternatively, polyclonal antibodies may be isolated from a mammal which has been previously infected with HCV.

Monoclonal antibodies directed against HCV epitopes can also be

10   readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980);

15   Hammerling et al. (1981); Kennett et al. (1980); see also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against HCV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, which are directed

20   against HCV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies.

Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is

25   desired. See, for example, Nisonoff, A., et al. (1981) and Dreesman et al. (1985). Techniques for raising anti-idiotype antibodies are known in the art. See, for example, Grzych (1985), MacNamara et al. (1984), and Vytdehaag et al. (1985). These anti-idiotype antibodies may also be useful for treatment, vaccination and/or diagnosis of NANBH, as well as for an elucidation of the

30   immunogenic regions of HCV antigens.

II.H. Immunoassay and Diagnostic Kits

Both the polypeptides and the antibodies of the present invention are useful in immunoassays to detect presence of HCV antibodies, or the presence of

the virus and/or HCV polypeptides (or epitopes), in, for example, biological samples. Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. The immunoassay will utilize at least one viral epitope derived from HCV. In one embodiment, the immunoassay uses a

5   combination of viral epitopes derived from HCV. These epitopes may be derived from the same or from different viral polypeptides, and may be in separate recombinant or natural polypeptides, or together in the same recombinant polypeptides. An immunoassay may use, for example, a monoclonal antibody directed towards a viral epitope(s), a combination of monoclonal antibodies directed

10   towards epitopes of one viral antigen, monoclonal antibodies directed towards epitopes of different viral antigens, polyclonal antibodies directed towards the same viral antigen, or polyclonal antibodies directed towards different viral antigens. Protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid sup-

15   ports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as

20   ELISA assays (described below).

Typically, an immunoassay for an anti-HCV antibody(s) will involve selecting and preparing the test sample suspected of containing the antibodies, such as a biological sample, then incubating it with an antigenic (*i.e.*, epitope-containing) HCV polypeptide(s) under conditions that allow antigen-antibody

25   complexes to form, and then detecting the formation of such complexes. Suitable incubation conditions are well known in the art. The immunoassay may be, without limitations, in a heterogenous or in a homogeneous format, and of a standard or competitive type.

In a heterogeneous format, the polypeptide is typically bound to a

30   solid support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (*e.g.*, in membrane or microtiter well form), polyvinyl chloride (*e.g.*, in sheets or microtiter wells), polystyrene latex (*e.g.*, in beads or microtiter plates,

polyvinylidine fluoride (known as Immulon), diazotized paper, nylon membranes, activated beads, and Protein A beads. For example, Dynatech Immulon 1 or Immulon® 2 microtiter plates or 0.25 inch polystyrene beads (Precision Plastic Ball) can be used in the heterogeneous format. The solid support containing the antigenic polypeptide is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are known in the art.

In a homogeneous format, the test sample is incubated with antigen in solution. For example, it may be under conditions that will precipitate any antigen-antibody complexes which are formed. Both standard and competitive formats for these assays are known in the art.

In a standard format, the amount of HCV antibodies forming the antibody-antigen complex is directly monitored. This may be accomplished by determining whether labeled anti-xenogenic (*e.g.*, anti-human) antibodies which recognize an epitope on anti-HCV antibodies will bind due to complex formation. In a competitive format, the amount of HCV antibodies in the sample is deduced by monitoring the competitive effect on the binding of a known amount of labeled antibody (or other competing ligand) in the complex.

Complexes formed comprising anti-HCV antibody (or, in the case of competitive assays, the amount of competing antibody) are detected by any of a number of known techniques, depending on the format. For example, unlabeled HCV antibodies in the complex may be detected using a conjugate of antixenogeneic Ig complexed with a label, (*e.g.*, an enzyme label).

In immunoassays where HCV polypeptides are the analyte, the test sample, typically a biological sample, is incubated with anti-HCV antibodies under conditions that allow the formation of antigen-antibody complexes. Various formats can be employed. For example, a "sandwich assay" may be employed, where antibody bound to a solid support is incubated with the test sample; washed; incubated with a second, labeled antibody to the analyte, and the support is washed again. Analyte is detected by determining if the second antibody is bound to the support. In a competitive format, which can be either heterogeneous or homogeneous, a test sample is usually incubated with antibody and a labeled,

competing antigen is also incubated, either sequentially or simultaneously. These and other formats are well known in the art.

Efficient detection systems for HCV infection may include the use of panels of epitopes, as described above. The epitopes in the panel may be constructed into one or multiple polypeptides. The assays for the varying epitopes may be sequential or simultaneous.

The enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration. Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing HCV epitopes or antibodies directed against HCV epitopes in suitable containers, along with the remaining reagents and materials required for the conduct of the assay, as well as a suitable set of assay instructions.

III. General Methods

The general techniques used in the practice of the present invention can be found in, for example, the references cited herein, particularly EPO Pub. Nos. 318,216 and 388,232, as well as the references in the bibliography, which are incorporated herein by reference.

IV.   Examples

Described below are examples of the present invention which are provided only for illustrative purposes, and not to limit the scope of the present invention. In light of the present disclosure, numerous embodiments within the

5   scope of the claims will be apparent to those of ordinary skill in the art.

IV.A.   Epitope Mapping of HCV Genome

The following example is the result of an epitope mapping experiment conducted on the HCV1 polyprotein sequence shown in Fig. 1. As

10   shown in Fig. 3-11, there are heterogeneities among HCV isolates, indicating that these amino acid substitutions can be made in the octamers described below. In addition to substitutions with amino acids from the corresponding location in other HCV isolates, substitutions with synthetic analogs of the particular amino acids or conservative substitutions based on charge, etc. (particularly when the substitution

15   does not destroy antibody binding) are within the scope of the invention.

IV.A.1.   Synthesis of Overlapping Peptides

Polyethylene pins arranged on a block in an 8×12 array (Coselco Mimetopes, Victoria, Australia) were prepared by placing the pins in a bath (20%

20   v/v piperidine in dimethylformamide (DMF)) for 30 minutes at room temperature. The pins were then removed, washed in DMF for 5 min, then washed in methanol four times (2 min/wash). The pins were allowed to air dry for at least 10 min, then washed a final time in DMF (5 min). 1-Hydroxybenzotriazole (HOBt, 367 mg) was dissolved in DMF (80 mL) for use in coupling Fmoc-protected amino

25   acids: Fmoc-L-Ala-OPfp, Fmoc-L-Cys(Trt)-OPfp, Fmoc-L-Asp(O-tBu)-OPfp, Fmoc-L-Glu(O-tBu)-OPfp, Fmoc-L-Phe-OPfp, Fmoc-Gly-OPfp, Fmoc-L-His(Boc)-OPfp, Fmoc-L-Ile-OPfp, Fmoc-L-Lys(Boc)-OPfp, Fmoc-L-Leu-OPfp, Fmoc-L-Met-OPfp, Fmoc-L-Asn-OPfp, Fmoc-L-Pro-OPfp, Fmoc-L-Gln-OPfp, Fmoc-L-Arg(Mtr)-OPfp, Fmoc-L-Ser(t-Bu)-ODhbt, Fmoc-L-Thr(t-Bu)-ODhbt, Fmoc-L-Val-

30   OPfp, and Fmoc-L-Tyr-OPfp.

The protected amino acids were placed in microtiter plate wells with HOBt, and the pin block placed over the plate, immersing the pins in the wells. The assembly was then sealed in a plastic bag and allowed to react at 25°C for 18

hours to couple the first amino acids to the pins. The block was then removed, and the pins washed with DMF (2 min), MeOH (4 × 2 min), and again with DMF (2 min) to clean and deprotect the bound amino acids. The procedure was repeated for each additional amino acid coupled, until all octamers had been prepared.

The free N-termini were then acetylated to compensate for the free amide, as most of the epitopes are not found at the N-terminus and thus would not have the associated positive charge. Acetylation was accomplished by filling the wells of a microtiter plate with DMF/acetic anhydride/triethylamine (5:2:1 v/v/v) and allowing the pins to react in the wells for 90 min at 20°C. The pins were then washed with DMF (2 min) and MeOH (4 × 2 min), and air dried for at least 10 min.

The side chain protecting groups were removed by treating the pins with trifluoroacetic acid/phenol/dithioethane (95:2.5:2.5, v/v/v) in polypropylene bags for 4 hours at room temperature. The pins were then washed in dichloromethane (2 × 2 min), 5% di-isopropylethylamine/dichloromethane (2 × 5 min), dichloromethane (5 min), and air-dried for at least 10 min. The pins were then washed in water (2 min), MeOH (18 hours), dried *in vacuo*, and stored in sealed plastic bags over silica gel.

IV.A.2.   Assay of Peptides

Octamer-bearing pins prepared as above were first treated by sonicating for 30 min in a disruption buffer (1% sodium dodecylsulfate, 0.1% 2-mercaptoethanol, 0.1 M NaH_2PO_4) at 60°C. The pins were then immersed several times in water (60°C), followed by boiling MeOH (2 min), and allowed to air dry. The pins were then precoated for 1 hour at 25°C in microtiter wells containing 200 μL blocking buffer (1% ovalbumin, 1% BSA, 0.1% Tween®, and 0.05% NaN_3 in PBS), with agitation. The pins were then immersed in microtiter wells containing 175 μL antisera obtained from human patients diagnosed as having HCV and allowed to incubate at 4°C overnight. The pins were assayed against antisera from three individual patients. Specimen #PAA 3663-s ("A") exhibited strong reaction to HCV Western blots, HCV competitive ELISA, HCV ELISA to clone C100-3 (at 1:1000 dilution), and RIBA responses of >4+ to C100, 5-1-1,

and C33c (C22 not done). (Antigen/clone names are per EPO Pub. Nos. 318,216 and 388,232, as well as those described in the literature regarding HCV immunoassays available from Ortho Diagnostics Systems, Inc.) Neat plasma was diluted 1:500 in blocking buffer. Specimen #PAA 33028 ("B") exhibited strong

5 reaction to HCV Western blots, HCV competitive ELISA, HCV ELISA to clone C100-3 (at 1:500 dilution), and RIBA responses of >4+ to C100, 5-1-1, C33C and C22. Polyclonal antisera was partially purified by passage through a protein A column, and was used at a dilution of 1:200 in blocking buffer. Specimen #PAA s32931 ("C") exhibited moderate reaction to HCV Western blots (3+),

10 HCV competitive ELISA, HCV ELISA to clone C100-3 (at 1:64 dilution), and RIBA responses of 3+ and 4+ to C100 and 5-1-1, respectively (C33c and C22 not done). Polyclonal antisera was partially purified by passage through a protein A column, and was used at a dilution of 1:500 in blocking buffer.

The pins were washed in PBS/Tween® 20 (4 × 10 min) at room

15 temperature, then incubated in microtiter wells containing horseradish peroxidase-labeled goat anti-Human Ig antisera (175 μL, 1:2000 dilution in blocking buffer without NaN$_3$) for 1 hour at 25°C with agitation. The antihuman antisera is specific for human Ig light and heavy chains, and reacts with both IgG and IgM classes. The pins were again washed in PBS/Tween® 20 (4 × 10 min) at room

20 temperature. Substrate solution was prepared by diluting NaH$_2$PO$_4$ (1 M, 200 mL) and citric acid (1 M, 160 mL) to 2 L with distilled water, adjusting the pH to 4.0. Azino-di-3-ethylbenzthiazodinsulfonate (ABTS, 50 mg) and hydrogen peroxide (0.3 μL/mL) was added to 100 mL of buffer immediately prior to use to complete the substrate solution. The substrate solution (150 μL) was added to each well of

25 a microtiter plate, and the pins immersed in the wells and incubated at 25°C in the dark. After color developed, the reactions were halted by removing the pins, and absorbance of the solutions read at 405 nm.

The octamers listed below were immunoreactive with anti-HCV antisera. Peptides reacting with all three antisera are listed as epitopes, while pep-

30 tides reacting with only one or two antisera are listed as weak epitopes (indicated by " − "). Particularly strong epitopes are labeled with letters rather than numbers (e.g., EpAA).

AA#	Sequence		AA#	Sequence
23	KEPGGGQA		87	GNEGCGWA
24	EPGGGQAV	Ep2		
25	PGGGQAVG		88	NEGCGWAG
26	GGGQAVGG		89	EGCGWAGW
27	GGQAVGGV		90	GCGWAGWL
28	GQAVGGVY		91	CGWAGWLL
29	QAVGGVYL ~		92	GWAGWLLS
30	AVGGVYLL		93	WAGWLLSP
31	VGGVYLLP		94	AGWLLSPR
32	GGVYLLPR		95	GWLLSPRG
33	GVYLLPRR		96	WLLSPRGS
34	VYLLPRRG		97	LLSPRGSR
35	YLLPRRGP		98	LSPRGSRP
36	LLPRRGPR		99	SPRGSRPS
66	PKARRPEG Ep1		100	PRGSRPSW Ep3
67	KARRPEGR		101	RGSRPSWG
68	ARRPEGRT		102	GSRPSWGP
69	RRPEGRTW		103	SRPSWGPT
70	RPEGRTWA			
71	PEGRTWAQ		186	TVPASAYQ Ep4
72	EGRTWAQP		187	VPASAYQV
73	GRTWAQPG EpA		188	PASAYQVR
74	RTWAQPGY		189	ASAYQVRN
75	TWAQPGYP		190	SAYQVRNS
76	WAQPGYPW		191	AYQVRNST
77	AQPGYPWP			
78	QPGYPWPL		206	DCPNSSIV ~
79	PGYPWPLG			
80	GYPWPLYG		223	TPGCVPCV ~
81	YPWPLYGN			
82	PWPLYGNE		232	EGNASRCW Ep5
83	WPLYGNEG			
84	PLYGNEGC		256	TQLRRHID ~
85	LYGNEGCG			
86	YGNEGCGW		286	LVGQLFTF ~
			297	RHWTTQGC Ep6
			298	HWTTQGCN
			299	WTTQGCNC
			321	MMMNWSPI ~
			347	DMIAGAHW Ep7
			357	LAGIAYFS Ep8

413	LINTNGSW EpB		594	YSRCGSGP F$_A$		
414	INTNGSWH		595	SRCGSGPW		
			596	RCGSGPWL		
432	SLNTGWLA ~		597	CGSGPWLT		
			598	GSGPWLTP		
465	FDQGWGPI EpC		599	SGPWLTPR		
466	DQGWGPIS		600	GPWLTPRC		
467	QGWGPISY		601	PWLTPRCL		
468	GWGPISYA		602	WLTPRCLV		
469	WGPISYAN		603	LTPRCLVD EpF		
470	GPISYANG		604	TPRCLVDY		
471	PISYANGS		605	PRCLVDYP		
			606	RCLVDYPY		
480	PDQRPYCW EpD		607	CLVDYPYR		
481	DQRPYCWH		608	LVDYPYRL		
482	QRPYCWHY		609	VDYPYRLW		
483	RPYCWHYP		610	DYPYRLWH		
484	PYCWHYPP		611	YPYRLWHY F$_B$		
			612	PYRLWHYP		
500	KSVCGPVY Ep9		613	YRLWHYPC		
501	SVCGPVYC					
502	VCGPVYCE		641	EAACNWTR		
		Ep12				
521	RSGAPTYS Ep10					
			662	LSPLLLII Ep13		
540	NNTRPPLG EpE		663	SPLLLIII		
541	NTRPPLGN		664	PLLLIIIQ		
542	TRPPLGNW		665	LLLIIIQW		
543	RPPLGNWF					
544	PPLGNWFG		685	LSTGLIHL ~		
545	PLGNWFGC					
546	LGNWFGCT		705	VGSSIASW ~		
547	GNWFGCTW		706	GSSIASWA		
548	NWFGCTWM					
549	WFGCTWMN		729	ARVCSCIW Ep14		
			782	WVPGAVYT		
579	LHCPTDCF ~					
		EpG	783	VPGAVYTF		
			784	PGAVYTFY		
			785	GAVYTFYG		
			786	AVYTFYGM		
			787	VYTFYGMW		
			788	YTFYGMWP		
			789	TFYGMWPL		
			801	LALPQRAY ~		

	851	RVEAQLHV EpH		1384	VIKGGRHL Ep20
	852	VEAQLHVW			
	853	EAQLHVWI		1410	LGINAVAY Ep21
	854	AQLHVWIP		1411	GINAVAYY
5	855	QLHVWIPP			
				1454	CNTCVIQT ~
	893	LAVFGPLW ~			
				1492	GRGKPGIY Ep22
	916	QGLLRFCA ~		1493	RGKPGIYR
10					
	928	MIGGHYVQ Ep15		1532	PAETTVRL Ep23
				1533	AETTVRLR
	946	TGTYVYNH EpI		1534	ETTVRLRA
				1535	TTVRLRAY
15	952	NHLTPLRD EpJ			
	953	HLTPLRDW		1560	GVFIGLIH Ep24
	954	LTPLRDWA		1561	VFIGLIHI
	1026	LAPITAYA ~		1581	ENLPYLVA
20			Ep25		
	1072	TCINGVCW EpK		1567	NLPYLVAY
				1568	LPYLVAYQ
	1109	LVGWPAPQ Ep16		1569	PYLVAYQA
				1570	YLVAYQAT
25	1171	CPAGHAVG Ep17		1571	LVAYQATV
	1113	PAGHAVGI		1572	VAYQATVC
	1114	AGHAVGIF		1573	AYQATVCA
	1115	GHAVGIFR		1574	YQATVCAR
	1116	HAVGIFRA		1575	QATVCARQ
30	1117	AVGIFRAA		1576	ATVCARQA
				1577	TVCARQAP
	1218	VVPQSEQV EpL			
				1601	PPSWDQMW
	1240	VPAAYAAQ ~	Ep26		
35				1602	PSWDQMWK
	1260	ATLGFGAY ~		1603	SWDQMWKC
				1604	WDQMWKCL
	1280	GVRITTGS Ep18		1605	DQMWKCLI
	1281	VRITTGSP		1606	QMWKCLIR
40	1282	RITTGSPI		1607	MWKCLIRL
	1283	ITTGSPIT			
	1284	TTGSPITY		1615	KPTLHGPI Ep27
	1285	TGSPITYG		1616	PTLHGPIP
				1617	TLHGPIPL
45	1322	DATSILGI ~		1618	LHGPIPLL
				1619	HGPIPLLY
	1338	TAGARLVV ~		1620	GPIPLLYR
	1371	GEIPFYGK Ep19		1655	VVTSTWVL ~

	1694	IIPDREVL EpM
	1695	IPDREVLY
	1710	ECSQHLPY EpN
5	1711	CSQHLPYI
	1712	SQHLPYIE
	1728	EKQKALGL EpO
	1729	KQKALGLL
10		
	1758	EIEWAKLM EpP
	1759	IEWAKLMW
	1760	EWAKLMWN
	1761	WAKLMWNE
15	1762	AKLMWNEI
	1781	LPGNPAIA ~
	1808	LFNILGGW Ep28
20		
	1821	AAPGAATA ~
	1851	ILAGYGAG Ep29
25	1880	VNLLPAIL ~
	1908	PGEGAVQW EpG
	1909	GEGAVQWM
	1910	EGAVQWMN
30	1911	GAVQWMNR
	1912	AVQWMNRL
	1913	VQWMNRLI
	1925	RGNHVSPI EpR
35		
	1940	AAARVTAI Ep30
	1941	AARVTAIL
	1942	ARVTAILS
	1943	RVTAILSS
40	1944	VTAILSSL
	1945	TAILSSLV
	1946	AILSSLVT
	1947	ILSSLVTQ
	1948	LSSLVTQL
45	1949	SSLVTQLL
	1950	SLVTQLLR
	1951	LVTQLLRR

1966	SECTIPCS EpS
1967	ECTIPCSG
1968	CTIPCSGS
1969	TIPCSGSW
1999	LMPQLPGI EpT
2000	MPQLPGIP
2001	PQLPGIPE
2002	QLPGIPEV
2003	LPGIPEVS
2004	PGIPEVSC
2005	GIPEVSCQ
2006	IPEVSCQR
2007	PEVSCQRG
2008	EVSCQRGY
2009	VSCQRGYK
2010	SCQRGYKG
2011	CQRGYKGV
2012	QRGYKGVW
2013	RGYKGVWR
2014	GYKGVWRG
2024	IMHTRCHC

Ep31

2048	VGPRICRN EpU
2049	GPRICRNY
2050	PRICRNYW
2051	RICRNYWS
2052	ICRNYWSG
2053	CRNYWSGT
2054	RNYWSGTE
2055	NYWSGTEP
2056	YWSGTEPI
2057	WSGTEPIN
2071	TPLPAPNY Ep32
2088	EEYVIRQV EpV
2089	EYVIRQVG
2090	YVIRQVGD
2091	VIRQVGDF
2092	IRQVGDFH
2093	RQVGDFHY
2108	DNLKCPCQ ~

	2122	EIELDGVR EpW
	2123	IELDGVRL
	2124	ELDGVRLH
	2125	LDGVRLHR
5	2126	DGVRLHRF
	2127	GVRLHRFA
	2128	VRLHRFAP
	2129	RLHRFAPP
	2130	LHRFAPPC
10	2131	HRFAPPCK
	2132	RFAPPCKP
	2133	FAPPCKPL
	2134	APPCKPLL
	2135	PPCKPLLR
15	2136	PCKPLLRE
	2137	CKPLLREE
	2138	KPLLREEV
	2139	PLLREEVS
	2140	LLREEVSF EpX
20	2141	LREEVSFR
	2142	REEVSFRV
	2143	EEVSFRVG
	2144	EVSFRVGL
	2145	VSFRVGLH
25	2146	SFRVGLHE
	2147	FRVGLHEY Ep38
	2148	RVGLHEYP
	2165	EPEPDVAV ~ EpAA
30		
	2187	GRRLARGS ~
	2226	LIEANLLW EpY
	2227	IEANLLWR
35	2228	EANLLWRQ
	2229	ANLLWRQE
	2230	NLLWRQEM
	2231	LLWRQEMG
	2232	LWRQEMGG
40		
	2244	VESENKVV Ep33
	2245	ESENKVVI
	2246	SENKVVIL
	2247	ENKVVILD
45	2248	NKVVILDS
	2249	KVVILDSF
	2250	VVILDSFD
	2267	EISVPAEI Ep34

2280	REAQALPV EpZ
2281	EAQALPVW
2282	AQALPVWA
2283	QALPVWAR
2284	ALPVWARP
2285	LPVWARPD
2286	PVWARPDY
2287	VWARPDYN
2288	WARPDYNP
2289	ARPDYNPP
2290	RPDYNPPL
2325	PPPRKKRT Ep35
2326	PPRKKRTV
2327	PRKKRTVV
2345	AELASRSE Ep36
2346	ELASRSEG
2347	LASRSEGS
2348	ASRSEGSS
2349	SRSEGSSS
2382	AESYSSMP Ep37
2401	SDGSWSTV
2417	VVCCSMSY
2418	VCCSMSYW
2419	CCSMSYWI
2420	CSMSYWIG
2421	SMSYWIGA
2422	MSYWIGAL

	2439	QKLPINAL EpBB		2602	LPLAVMGS
	2440	KLPINALS	EpDD		
	2441	LPINALSN		2603	PLAVMGSS
	2442	PINALSNS		2604	LAVMGSSY
5	2443	INALSNSL		2605	AVMGSSYG
	2444	NALSNSLL		2606	VMGSSYGE
	2445	ALSNSLLR		2607	MGSSYGEQ
	2446	LSNSLLRH		2608	GSSYGEQR
	2447	SNSLLRHH		2609	SSYGEQRV
10	2448	NSLLRHHN		2610	SYGEQRVE
	2449	SLLRHHNL		2611	YGEQRVEE
	2450	LLRHHNLV		2612	GEQRVEEL
	2451	LRHHNLVY			
	2452	RHHNLVYS		2632	KTPMGFSY
15	2453	HHNLVYST	Ep41		
	2454	HNLVYSTI		2633	TPMGFSYD
	2455	NLVYSTIS		2634	PMGFSYDT
	2456	LVYSTISR		2635	MGFSYDTR
				2636	GFSYDTRC
20	2469	QKKVTFDR Ep39		2637	FSYDTRCE
	2470	KKVTFDRL		2638	SYDTRCED
	2471	KVTFDRLQ			
	2472	VTFDRLQV		2660	YQCCDLDP ~
	2473	TFDRLQVL			
25	2474	FDRLQVLD		2676	LTERLYVG
	2475	DRLQVLDS	EpEE		
	2476	RLQVLDSH		2677	TERLYVGG
				2678	ERLYVGGP
	2495	ASKVKANL ~		2679	RLYVGGPL
30					
	2533	RKAVTHIN EpCC		2688	NSRGENCG
	2534	KAVTHINS	EpFF		
				2689	SRGENCGY
	2573	GRKPARLI Ep40		2690	RGENCGYR
35	2574	RKPARLIV		2691	GENCGYRR
	2575	KPARLIVF		2692	ENCGYRRC
	2576	PARLIVFP		2693	NCGYRRCR
	2577	ARLIVFPD			
	2578	RLIVFPDL		2707	TSCGNTLI Ep42
40					
				2721	AACRAAGL ~
				2757	AFTEAMTR
45			Ep43		
				2758	FTEAMTRY
				2759	TEAMTRYS
				2760	EAMTRYSA
				2761	AMTRYSAP
				2762	MTRYSAPP

		2779	DLELIISC Ep44	2878	DLPPIIQR Ep47
		2780	LELIISCS	2879	LPPIIQRL
				2880	PPIIQRLH
		2794	HDGAGKRV	2881	PIIQRLHG
5	Ep45			2882	IIQRLHGL
		2795	DGAGKRVY	2883	IQRLHGLS
		2796	GAGKRVYY	2884	QRLHGLSA
		2797	AGKRVYYL	2885	RLHGLSAF
		2798	GKRVYYLT	2886	LHGLSAFS
10		2799	KRVYYLTR	2887	HGLSAFSL
		2800	RVYYLTRD	2888	GLSAFSLH
		2801	VYYLTRDP	2889	LSAFSLHS
		2802	YYLTRDPT	2890	SAFSLHSY
				2891	AFSLHSYS
15		2817	WETARHTP	2892	FSLHSYSP
	EpGG			2893	SLHSYSPG
		2818	ETARHTPV	2894	LHSYSPGE
		2819	TARHTPVN	2895	HSYSPGEI
		2820	ARHTPVNS		
20		2821	RHTPVNSW		
		2822	HTPVNSWL		
		2823	TPVNSWLG		
		2824	PVNSWLGN		
		2825	VNSWLGNI		
25		2826	NSWLGNII		
		2827	SWLGNIIM		
		2828	WLGNIIME		
		2829	LGNIIMEA		
		2830	GNIIMEAP		
30		2831	NIIMEAPT		
		2832	IIMEAPTL		
		2833	IMEAPTLW		
		2834	MEAPTLWA		
		2835	EAPTLWAR		
35		2836	APTLWARM		
		2837	PTLWARMI		
		2838	TLWARMIL		
		2839	LWARMILM		
		2840	WARMILMT		
40		2841	ARMILMTH		
		2842	RMILMTHF		
		2843	MILMTHFE		
		2863	LDCEIYGA Ep46		
45		2864	DCEIYGAC		
		2865	CEIYGACY		
		2866	EIYGACYS		
		2867	IYGACYSI		

2912	LGVPPLRA Ep48
2913	GVPPLRAW
2914	VPPLRAWR

5	2938	AAICGKYL Ep49
	2939	AICGKYLE
	2940	ICGKYLEN

		2966	DLSGWETA
10	EpHH		

2984	VSHARPRW EpII

15   IV.B.   <u>Differential Assay</u>

The following assay was performed to distinguish early antigens
from later antigens. Antibodies to the early antigens may be detected, and used to
diagnose HCV infection more quickly.

Serial bleeds were obtained from a human patient presenting with
20   elevated ALT, but negative for anti-C100-3 antibody. The five bleeds obtained
prior to complete seroconversion (C100-3 positive) were pooled and used in the
assay at a dilution of 1:2000. The assay was conducted as described in Section
IV.A. above. However, one duplicate set of pins was incubated with horseradish
peroxidase-labeled goat anti-Human IgG specific antisera, while the other set was
25   incubated with horseradish peroxidase-labeled goat anti-Human IgM specific anti-
sera. Epitopes immunoreactive with IgM antibodies are early epitopes.

The results indicated that most early epitopes are found in the
region extending from about amino acid 480 to about amino acid 650. Particularly
strong IgM epitopes were octamers beginning with amino acid nos. 506, 510, 523,
30   553, 562, 580, and the region from 590 to 620. Assays which employ antigens
bearing epitopes from this region will permit diagnosis of HCV infection at an
early point than assays employing other antigens.

We have additionally tested serial plasma specimens taken from five
patients with open heart post-transfusion NANB hepatitis, with studies followed
35   for 3-12 years. Initial bleed dates were less than one week apart. Each specimen
was tested for IgG and IgM by EIA against one core antigen (C22), two envelope
antigens (E1 and E2), and three nonstructural region antigens (C33c, C100, and

NS5). We found that the IgM response to C22 and C33c preceeded the IgG response for those antigens. NS-5 also induced an IgM response, but this response did not preceed the IgG response for that antigen. Thus, one can prepare assays capable of determining very early stages of infection by utilizing epitopes

5 derived from the C22 and C33c regions and assaying for IgM binding. Antibodies to the C33c region persisted for the longest period of time, suggesting that diagnostic assays directed toward C33c should be the most reliable.

IV.C. Sequence Variations in HCV Isolates from Different Individuals

10 Isolates of HCV which contain sequences which deviate from CDC/HCV1 were identified in human individuals, some of whom were serologically positive for anti-C100-3 antibodies (EC10 was antibody negative). Identification of these new isolates was accomplished by cloning and sequencing segments of the HCV genome which had been amplified by the PCR technique

15 using CDC/HC1 sequences. The method utilizes primers and probes based upon the HCV cDNA sequences described herein. The first step in the method is the synthesis of a cDNA to either the HCV genome, or its replicative intermediate, using reverse transcriptase. After synthesis of the HCV cDNA, and prior to amplification, the RNA in the sample is degraded by techniques known in the art.

20 A designated segment of the HCV cDNA is then amplified by the use of the appropriate primers. The amplified sequences are cloned, and clones containing the amplified sequences are detected by a probe which is complementary to a sequence lying between the primers, but which does not overlap the primers.

25 IV.C.1. HCV Isolates Isolated from Humans in the U.S.

Blood samples which were used as a source of HCV virions were obtained from the American Red Cross in Charlotte, North Carolina, and from the Community Blood Center of Kansas, Kansas City, Missouri. The samples were screened for antibodies to the HCV C100-3 antigen using an ELISA assay and

30 subjected to supplemental Western blot analysis using a polyclonal goat anti-human HRP to measure anti-HCV antibodies. Two samples, #23 and #27, from the American Red Cross and from the Community Blood Center of Kansas, respectively, were determined to be HCV positive by these assays.

Viral particles present in the serum of these samples were isolated by ultracentrifugation under the conditions described by Bradley *et al.* (1985). RNA was extracted from the particles by digestion with proteinase K and SDS at final concentrations of 10 μg/mL proteinase K, and 0.1% SDS; digestion was for

5   1 hour at 37°C. Viral RNA was further purified by extraction with chloroform-phenol.

HCV RNA in the preparation of RNA was reverse transcribed into cDNA. After both strands of the cDNA were synthesized, the resulting cDNA was then amplified by the PCR method. The HCV cDNAs in three clones derived

10 from each HCV isolate, were subjected to sequence analysis. Analysis was essentially by the method described in Chen and Seeburg (1985).

Consensus sequences of the clones derived from HCV in samples 23 and 27 are shown in Fig. 3 and Fig. 4, respectively. The variable sequences are also shown in these figures, as are the amino acids encoded in the consensus

15 sequences.

Figures 5 and 6 show comparisons of the aligned positive strand nucleotide sequences (Fig. 5) and putative amino acid sequences (Fig. 6) of samples 23, 27, and HCV1. The amino acid sequence of HCV1 in Fig. 6 represents amino acid numbers 129-467 of the HCV polyprotein encoded by the

20 large ORF in the HCV genomic RNA. An examination of Figs. 5 and 6 show that there are variations in the sequences of the three isolated clones. The sequence variations at the nucleotide level and the amino acid level are summarized in the table immediately below. In the table, the polypeptides designated S and NS1 represent amino acid numbers 130 to ~380, and 380 to ~470, respectively, as

25 those domains were previously known. The numbering is from the putative initiator methionine. The terminology S and NS1 is based upon the positioning of the sequences encoding the polypeptides using the Flavivirus model. As discussed above, however, recent evidence suggests that there is not total correlation between HCV and the Flaviviruses with regard to viral polypeptide domains,

30 particularly in the putative E/NS1 domains. Indeed, HCV polypeptides and their coding domains may exhibit substantial deviation from the Flavivirus model.

TABLE:

	Sequence Homology					
	Nucleotide Encoding			Amino Acid Encoded		
	overall	S	NS1	overall	S	NS1
	%	%	%	%	%	%
HCV1/HCV23	93	95	91	92	95	87
HCV1/HCV27	89	93	84	89	95	82
HCV23/HCV27	89	93	85	90	93	84

Although there are variations in the newly isolated HCV sequences, the cloned sequences from samples 23 and 27 (called HCV23 and HCV27) each contain 1019 nucleotides, indicating a lack of deletion and addition mutants in this region in the selected clones. The sequences in Figs. 5 and 6 also show that the isolated sequences are not rearranged in this region.

A comparison of the consensus sequences for HCV1 and for the other isolates of HCV is summarized in the Table, *supra*. The sequence variations between the chimpanzee isolate HCV1, and the HCVs isolated from humans are about the same as that seen between the HCVs of human origin.

It is of interest that the sequence variations in two of the putative domains is not uniform. The sequence in a putative S region appears to be relatively constant, and randomly scattered throughout the region. In contrast, a putative NS1 region has a higher degree of variability than the overall sequence, and the variation appears to be in a hypervariable pocket of about 28 amino acids which is located about 70 amino acids downstream from the putative N-terminus of the putative polyprotein.

Although it may be argued that the detected variations were introduced during the amplification process, it is unlikely that all of the variations are from this result. It has been estimated that Taq polymerase introduces errors into a sequence at approximately one base per 10 kilobases of DNA template per cycle (Saiki *et al.* (1988)). Based upon this estimate, up to 7 errors may have been introduced during the PCR amplification of the 1019 bp DNA fragment. However, the three subclones of HCV-23 and HCV-27 yielded 29 and 14 base variations, respectively. The following suggest that these variations are naturally

occurring. About 60% of the base changes are silent mutations which do not change the amino acid sequence. Variations introduced by the Taq polymerase during PCR amplification would be expected to occur randomly; however, the results show that the variant sequences are clustered in at least one specific region.

IV.C.2. HCV Isolates from Humans in Italy and in the U.S.

Segments of HCV RNA present in different isolates were amplified by the HCV/cPCR method. These segments span a region of ~0.6 Kb to ~1.6 Kb downstream from the methionine encoding start codon of the putative HCV polyprotein. The isolates are from biological specimens obtained from HCV infected individuals. More specifically, isolate HCT #18 is from human plasma from an individual in the U.S.A., EC1 and EC10 are from a liver biopsy of an Italian patient, and Th is from a peripheral blood mononucleocyte fraction of an American patient. Comparable segments of HCV RNA have been isolated from a chimpanzee.

RNA was extracted from the human plasma specimens using phenol:CHCl$_3$:isoamyl alcohol extraction. Either 0.1 mL or 0.01 mL of plasma was diluted to a final volume of 1.0 mL, with a TENB/proteinase K/SDS solution (0.05 M Tris-HCL, pH 8.0, 0.001 M EDTA, 0.1 M NaCl, 1 mg/mL Proteinase K, and 0.5% SDS) containing 10 to 40 μg/mL polyadenylic acid, and incubated at 37°C for 60 minutes. After this proteinase K digestion, the resultant plasma fractions were deproteinized by extraction with TE (50 mM Tris-HCl, pH 8.0, 1 mM EDTA) saturated phenol, pH 6.5. The phenol phase was separated by centrifugation, and was reextracted with TENB containing 0.1% SDS. The resulting aqueous phases from each extraction were pooled, and extracted twice with an equal volume of phenol/chloroform/isoamyl alcohol [1:1(99:1)], and then twice with an equal volume of a 99:1 mixture of chloroform/isoamyl alcohol. Following phase separation by centrifugation, the aqueous phase was brought to a final concentration of 0.2 M Na acetate, and the nucleic acids were precipitated by the addition of two volumes of ethanol. The precipitated nucleic acids were recovered by ultracentrifugation in a SW 41 rotor at 38 K, for 60 minutes at 4°C or in a microfuge for 10 minutes at 10 K, 4°C.

RNA extracted from the liver biopsy was provided by Dr. F.

Bonino, Ospedale Maggiore di S. Giovanni Battista, Torino, Italy. The

mononucleocyte fraction was obtained by sedimentation of the individual's aliquot

of blood through Ficoll-Paque® (Pharmacia Corp), using the manufacturer's dir-

5   ections. Total RNA was extracted from the fraction using the guanidinium thio-

cyanate procedure described in Choo et al (1989).

Synthesis of HCV cDNA from the samples was accomplished using

reverse transcriptase. Following ethanol precipitation, the precipitated RNA or

nucleic acid fraction was dried, and resuspended in DEPC treated distilled water.

10 Secondary structures in the nucleic acids were disrupted by heating at 65°C for 10

minutes, and the samples were immediately cooled on ice. cDNA was synthesized

using 1 to 3 μg of total RNA from liver, or from nucleic acids (or RNA) extracted

from 10 to 100 μL of plasma. The synthesis utilized reverse transcriptase, and

was in a 25 μL reaction, using the protocol specified by the manufacturer, BRL.

15 All reaction mixtures for cDNA synthesis contained 23 units of the RNAase

inhibitor, Rnasin® (Fisher/Promega). Following cDNA synthesis, the reaction

mixtures were diluted with water, boiled for 10 minutes, and quickly chilled on

ice.

Each set of samples was subjected to two rounds of PCR

20 amplification. The primers for the reactions were selected to amplify regions

designated "EnvL" and EnvR". The "EnvL" region encompasses nucleotides 669-

1243, and putative amino acids 117 to 308; the "EnvR" region encompasses

nucleotides 1215-1629, and encodes putative amino acids 300-408 (the putative

amino acids are numbered starting from the putative methionine initiation codon).

25   The PCR reactions were performed essentially according to the

manufacturer's directions (Cetus-Perkin-Elmer), except for the addition of 1 μg of

RNase A. The reactions were carried out in a final volume of 100 μL . The PCR

was performed for 30 cycles, utilizing a regimen of 94°C (1 min), 37°C (2 min),

and 72°C (3 min), with a 7 minute extension at 72°C for the last cycle. The

30 samples were then extracted with phenol:CHCl$_3$, ethanol precipitated two times,

resuspended in 10 mM Tris HCl, pH 8.0, and concentrated using Centricon-30

(Amicon) filtration. This procedure efficiently removes oligonucleotides less than

30 nucleotides in size; thus, the primers from the first round of PCR amplification are removed.

The Centricon-30 concentrated samples were then subjected to a second round of PCR amplification. Amplification by PCR was for 35 cycles

5   utilizing a regimen of 94°C (1 min), 60°C (1 min), and 72°C (2 min), with a 7 minute extension at 72°C for the last cycle. The samples were then extracted with phenol:CHCl₃, precipitated two times, and digested with EcoRI. The PCR reaction products were analyzed by separation of the products by electrophoresis on 6% polyacrylamide gels. DNA of approximately the estimated size of the

10   expected PCR product was electroeluted from the gels, and subcloned into either a pGEM-4 plasmid vector or into λgt11. The expected product sizes for the EnvL and EnvR after the first round of amplification are 615 bp and 683 bp, respectively; after the second round of amplification the expected product sizes for EnvL and EnvR are 414 bp and 575 bp, respectively. The plasmids containing the

15   amplified products were used to transform host cells; the pGEM-4 plasmid was used to transform DH5-alpha, and λgt11 was used to transform C600 delta-HFL. Clones of the transformed cells which either hybridized to the appropriate HCV probes, or those which had inserts of the correct size were selected. The inserts were then cloned in M13 and sequenced. The probes for all of the HCV/cPCR

20   products consisted of ³²P labeled sections of HCV cDNA which had been prepared by PCR amplification.

Sequence information on variants in the EnvL region was obtained from 3 clones from HCT #18, 2 clones from TH, 3 clones from EC1, and from the HCV1 clones. A comparison of the composite nucleotide sequence of each

25   isolate derived from these clones is shown in Fig. 7. In the figure, each sequence is shown 5' to 3' for the sense strand for the EnvL region, and the sequences have been aligned. The vertical lines and capital letters indicate sequence homology, the absence of a line and an uncapitalized letter indicates a lack of homology. The sequences shown in the lines are as follows: line 1, Thorn; line 2, EC1; line 3,

30   HCT #18; line 4, HCV1.

Sequence information on variants in the EnvR region was obtained from two clones of EC10, and from HCV1 clones. The two EC10 clones differed by only one nucleotide. A comparison of the nucleotide sequences of EC10(clone

2) and a composite of the HCV1 sequences is shown in Fig. 8; each sequence is shown 5' to 3' for the sense strand of the EnvR region, and the sequences have been aligned. The double dots between the sequences indicate sequence homology.

5   A comparison of the amino acid sequences encoded in the EnvL (amino acids #117-308) and EnvR region (amino acids #300-438) for each of the isolates is shown in Fig. 9 and Fig. 10, respectively. Included in the Figures are sequences for the isolates JH23 and JH27, described above. Also indicated are sequences from a Japanese isolate; these sequences were provided by Dr. T. 10   Miyamura, Japan. In the figures, the amino acid sequence for the region is given in its entirety for HCV1, and the non-homologous amino acids in the various isolates are indicated.

As seen in Fig. 9, In the EnvL region there is overall about a 93% homology between HCV1 and the other isolates. HCT18, Th, and EC1 have 15   about a 97% homology with HCV1; JH23 and JH27 have about 96% and about 95% homology, respectively, with HCV1. Fig. 10 shows that the homologies in the EnvR region are significantly less than in the EnvL region; moreover, one subregion appears to be hypervariable (*i.e.*, from amino acid 383-405). This data is summarized in the Table immediately below.

20   Table: <u>Homology of EnvR Region</u>

Isolate	Percent Homology with HCV1	
	AA330-AA438	AA383-AA405
JH23(U.S.)	83	57
JH27(U.S.)	80	39
Japanese	73	48
EC10 (Italy)	84	48

VI.   <u>Industrial Applicability</u>

The epitopes identified herein can be used to make polypeptide 30   products as described above for applications such as the screening of blood for HCV infection, clinical HCV diagnosis, the generation of antibodies, and preparation of medicaments. Other applications are described above, and still others will be readily apparent to those of ordinary skill.

VII.   Bibliography

Barr *et al.*, Biotechniques (1986) 4:428.

Bradley *et al.*, Gastroenterology (1985) 88:773.

Botstein, Gene (1979) 8:17.

5   M.A. Brinton, (1986) in "The Viruses: The Togaviridae and Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.327-374.

Broach (1981) in: "Molecular Biology of the Yeast - Saccharomyces, Vol. 1, p.445, Cold Spring Harbor Press.

10   Broach *et al.*, Meth Enzymol (1983) 101:307.

Catty (1988), "Antibodies, Volume 1: A Practical Approach" (IRL Press).

Chaney *et al.*, Cell Mol Genet (1986) 12:237.

Chakrabarti *et al.*, Mol Cell Biol (1985) 5:3403.

Chang *et al.*, Nature (1977) 198:1056.

15   Chen and Seeburg, DNA (1985) 4:165.

Chirgwin *et al.*, Biochemistry (1979) 18:5294.

Chomczynski and Sacchi, Anal Biochem (1987) 162:156.

Choo *et al.*, Science (1989) 244:359.

Clewell *et al.*, Proc Natl Acad Sci USA (1969) 62:1159.

20   Clewell, J Bacteriol (1972) 110:667.

Cohen, Proc Natl Acad Sci USA (1972) 69:2110.

Cousens *et al.*, Gene (1987) 61:265.

De Boer *et al.*, Proc Natl Acad Sci USA (1983) 292:128.

Dreesman *et al.*, J Infect Dis (1985) 151:761.

25   S.M. Feinstone and J.H. Hoofnagle, New Engl J Med (1984) 311:185.

Felgner *et al.*, Proc Natl Acad Sci USA (1987) 84:7413.

Fields & Knipe (1986), "Fundamental Virology" (Raven Press, N.Y.).

Fiers *et al.*, Nature (1978) 273:113.

R.J. Gerety *et al.*, in "Viral Hepatitis and Liver Disease" (B.N. Vyas, J.L.

30   Dienstag, and J.H. Hoofnagle, eds)

Glennie *et al.*, Nature (1982) 295:712.

Gluzman, Cell (1981) 23:175.

Goeddel *et al.*, Nuc Acids Res (1980) 8:4057.

Graham and Van der Eb, Virology (1978) 52:546.

Grunstein and Hogness, Proc Natl Acad Sci USA (1975) 73:3961.

Grych et al., Nature (1985) 316:74.

Gubler and Hoffman Gene (1983) 25:263.

5   Hahn, Virology (1988) 162:167.

Hammerling et al., (1981), "Monoclonal Antibodies and T-Cell Hybridomas".

Han, Biochemistry (1987) 26:1617.

Helfman, Proc Natl Acad Sci USA (1983) 80:31.

Hess et al., J Adv Enzyme Reg (1968) 7:149.

10   Hinnen et al., Proc Natl Acad Sci USA (1978) 75:1929.

Hitzeman et al., J Biol Chem (1980) 255:2073.

Holland et al., Biochemistry (1978) 17:4900.

Holland, J Biol Chem (1981) 256:1385.

Holland and Holland, J Biol Chem (1980) 255:2596.

15   Hoopes et al., Nuc Acids Res (1981) 9:5493.

Houghton et al., Nuc Acids Res (1981) 9:247

T.V. Hunyh et al., (1985) in "DNA Cloning Techniques; A Practical Approach"

(D. Glover, Ed., IRL Press, Oxford, U.K.) pp. 49-78.

Immun Rev (1982) 62:185.

20   Ito et al., Agric Biol Chem (1984) 48:341.

Iwarson, British Medical J (1987) 295:946.

Kennett et al. (1980) "Monoclonal Antibodies".

Kniskern et al., Gene (1986) 46:135.

Kyte and Doolittle, J Mol Bio (1982) 157:105-132.

25   Laemmli, Nature (1970) 227:680.

Lee et al., Science (1988) 239:1288.

Luckow and Summers, Virol (1989) 17:31.

Mackett et al., J Virol (1984) 49:857.

T. Maniatis et al. (1982) "Molecular Cloning; A Laboratory Manual" (Cold

30   Spring Harbor Press, Cold Spring Harbor, N.Y.).

Sambrook et al. (1989), "Molecular Cloning; A Laboratory Manual", Second

Edition (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).

Manning and Mocarski, Virol (1988) 167:477.

Mayer and Walker, eds. (1987), "Immunochemical Methods In Cell And Molecular Biology" (Academic Press, London).

Maxam et al., Meth Enzymol (1980) 65:499.

MacNamara et al., Science (1984) 226:1325.

5   Messing et al., Nuc Acids Res (1981) 9:309.

Messing, Meth Enzymol (1983) 101:20-37.

Michelle et al., Int. Symposium on Viral Hepatitis.

Monath (1986) in "The Viruses: The Togaviradae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum

10   Press), pp. 375-440.

Moss (1987) in "Gene Transfer Vectors For Mammalian Cells" (Miller and Calos, eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), p. 10.

Nagahuma et al., Anal Biochem (1984) 141:74.

Neurath et al., Science (1984) 224:392.

15   Nisonoff et al., Clin Immunol Immunopathol (1981) 21:397-406.

L.R. Overby, Curr Hepatol (1985) 5:49.

L.R. Overby, Curr Hepatol (1986) 6:65.

L.R. Overby, Curr Hepatol (1987) 7:35.

Pachl et al., J Virol (1987) 61:315.

20   Peleg, Nature (1969) 221:193.

Pfefferkorn and Shapiro (1974), in "Comprehensive Virology", Vol. 2 (Fraenkel-Conrat & Wagner, eds., Plenum, N.Y.) pp. 171-230.

A.M. Prince, Ann Rev Microbiol (1983) 37:217.

Rice et al., Science (1985) 229:726.

25   Rice et al. (1986) in "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.279-328.

Roehrig (1986) in "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum

30   Press)

Sadler et al., Gene (1980) 8:279.

Saiki et al., Nature (1986) 324:163.

Saiki et al., Science (1988) 239:487.

Sanger *et al.*, Proc Natl Acad Sci USA (1977) 74:5463.

Setlow, ed. (1988), "Genetic Engineering" Vol. 10, pp. 195-219 (Plenum Publishing Co., N.Y.

Schlesinger *et al.*, J Virol (1986) 60:1153.

M. Schreier *et al.*, (1980) "Hybridoma Techniques"

Scopes (1984), "Protein Purification, Principles and Practice", 2nd Ed., (Springer-Verlag, N.Y.).

Shimatake *et al.*, Nature (1981) 292:128.

Singh *et al.*, Nuc Acids Res (1983) 11:4049.

Sippel, Eur J Biochem (1973) 37:31.

Smith *et al.*, Mol Cell Biol (1983) 3:2156-2165.

Steimer *et al.*, J Virol (1986) 58:9.

Stollar (1980), in "The Togaviruses" (R.W. Schlesinger, ed.), pp. 584-622.

Stuve *et al.*, J Virol (1987) 61:326.

Sumiyoshi *et al.*, Virol (1987) 161:497.

Taylor *et al.*, Biochim Biophys Acta (1976) 442:324.

Towbin *et al.*, Proc Natl Acad Sci USA (1979) 76:4350.

Tsu and Herzenberg (1980), in "Selected Methods In Cellular Immunology" (W.H. Freeman and Co.) pp. 373-391.

Vytdehaag *et al.*, J Immunol (1985) 134:1225.

P. Valenzuela *et al.*, Nature (1982) 298:344.

P. Valenzuela *et al.*, (1984), in "Hepatitis B" (I. Millman *et al.*, ed, Plenum Press) pp. 225-236.

Warner, DNA (1984) 3:401.

Ward *et al.*, Nature (1989) 341:544.

Wu and Grossman, Meth Enzymol (1987), 154 "Recombinant DNA", Part E.

Wu, Meth Enzymol (1987), 155, "Recombinant DNA", part F.

Zoller, Nuc Acids Res (1982) 10:6487.

U.S. Patent No. 4,341,761   U.S. Patent No. 4,466,917

U.S. Patent No. 4,399,121   U.S. Patent No. 4,472,500

U.S. Patent No. 4,427,783   U.S. Patent No. 4,491,632

U.S. Patent No. 4,444,887   U.S. Patent No. 4,493,890

U.S. Patent No. 4,816,467

WHAT IS CLAIMED:

1.   A polypeptide comprising a truncated HCV sequence containing an HCV epitope of the formula

$$aa_x\text{-}aa_y$$

wherein aa denotes an amino acid;

x and y are integers such that $y\text{-}x \geq 6$;

$aa_x\text{-}aa_y$ indicates a portion of the amino acid sequence of Figure 1; and

x is selected from the group consisting of 23-34, 36, 66-79, 81-94, 96-98, 101-103, 186-189, 191, 206, 223, 232, 256, 286, 297-299, 321, 347, 357, 413, 414, 432, 465-471, 480-484, 501, 502, 521, 540-549, 579, 594-599, 601-613, 641, 662-665, 685, 705, 706, 729, 782-789, 801, 851-855, 893, 916, 928, 946, 952-954, 1026, 1072, 1109, 1112-1117, 1218, 1240, 1280-1285, 1322, 1338, 1371, 1384, 1410, 1411, 1454, 1492, 1493, 1532-1535, 1560, 1561, 1566-1568, 1571-1577, 1601-1607, 1615-1620, 1655, 1695, 1710-1712, 1728, 1729, 1758-1762, 1781, 1808, 1821, 1851, 1880, 1908-1913, 1925, 1940-1948, 1951, 1966-1969, 1999, 2001-2004, 2006-2014, 2024, 2048-2053, 2055-2057, 2071, 2088-2093, 2108, 2122-2148, 2165, 2187, 2226-2232, 2244-2249, 2267, 2281-2286, 2288, 2289, 2325-2327, 2346, 2347, 2349, 2382, 2401, 2417-2422, 2439-2444, 2446-2456, 2469, 2471-2476, 2495, 2533, 2534, 2573-2578, 2602-2604, 2606-2612, 2632-2638, 2660, 2676-2679, 2688-2693, 2707, 2721, 2757-2762, 2779, 2794, 2795, 2797-2799, 2801, 2802, 2817-2843, 2863-2867, 2878-2884, 2886-2895.

2.   The polypeptide of claim 1 which is about 100 amino acids or less in length.

3.   The polypeptide of claim 2 wherein $y\text{-}x \leq 50$.

4.   The polypeptide of claim 2 wherein $y\text{-}x \leq 20$.

5.   The polypeptide of claim 2 wherein $y\text{-}x < 10$.

6.   The polypeptide of claim 2 wherein x is selected from the group consisting of 506, 510, 523, 553, 562, 580, and 590-620.

7.   A polypeptide of about 100 amino acids or less comprising an
5   HCV epitope of the formula

$$aa_x\text{-}aa_y$$

wherein aa denotes an amino acid;

x and y are integers such that $y\text{-}x \geq 6$;

$aa_x\text{-}aa_y$ indicates a portion of the amino acid sequence of Figure 1; and
10   x is selected from the group consisting of 35 (where y is less than 45), 80 (where y is less than 90), 95 (where y is less than 110), 99 (where y is less than 120), 100 (where y is less than 150), 190 (where y is less than 210), 500 (where y is less than 550), 600 (where y is less than 625), 1260 (where y is less than 1280), 1569 (where y is less than 1931), 1570 (where y is less than 1590),
15   1694 (where y is less than 1735), 1949 (where y is less than 2124), 1950 (where y is less than 1985), 2000 (where y is less than 2050), 2005 (where y is less than 2025), 2054 (where y is less than 2223), 2250 (where y is less than 2330), 2287 (where y is less than 2385), 2290 (where y is less than 2310), 2345 (where y is less than 2375), 2348 (where y is less than 2464), 2445 (where y is less than
20   2475), 2470 (where y is less than 2490), 2605 (where y is less than 2620), 2780 (where y is less than 2830), 2796 (where y is less than 2886), 2800 (where y is less than 2850), and 2885 (where y is less than 2905).

8.   An immunoassay reagent comprising a polypeptide according to
25   any of claims 1-7.

9.   The immunoassay reagent of claim 8 wherein $y\text{-}x \leq 50$.

10.   The immunoassay reagent of claim 8 wherein x is selected from
30   the group consisting of 506, 510, 523, 553, 562, 580, and 590-620.

11. A method for detecting the presence of antibodies immunoreactive with Hepatitis C virus (HCV) proteins in a sample, said method comprising:

contacting an immobilized immunoassay reagent according to claim 8 with said sample; and

5   detecting antibodies bound to said reagent.

12.   A method for inducing an immunological response in a subject against HCV, said method comprising:

10   administering to said subject an effective amount of a polypeptide accoding to any one of claims 1-7.

13.   A composition for inducing an immunological response in a subject against HCV, said composition comprising an effective amount of a

15   polypeptide accoding to any one of claims 1-7.

14.   A monoclonal or polyclonal antibody composition wherein said antibodies bind the HCV epitope of a polypeptide according to claim 1.

20   15.   A method of making a polypeptide according to any of claims 1-7 wherein said polypeptide is prepared by recombinant expression or chemical synthesis.

FIG. 1 − 1

```
          R  T
MSTNPKPQKKNKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATR
KTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSP-100
RGSRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARA


                  T
LAHGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSTGL-200
YHVTNDCPNSSIVYEAADAILHTPGCVPCVREGNASRCWVAMTPTVATRD
GKLPATQLRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRRHWT-300


                  V
TQGCNCSIYPGHITGHRMAWDMMMNWSPTTALVMAQLLRIPQAILDMIAG
AHWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDAETHVTGGSAGHTVSGFV-400
SLLAPGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLAGLFYHHKFNSS
GCPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCWHYPPKPCGIVPAK-500
SVCGPVYCFTPSPVVVGTTDRSGAPTYSWGENDTDVFVLNNTRPPLGNWF
GCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSG-600


    I
PWLTPRCLVDYPYRLWHYPCTINYTIFKIRMYVGGVEHRLEAACNWTRGE
RCDLEDRDRSELSPLLLTTTQWQVLPCSFTTLPALSTGLIHLHQNIVDVQ-700
YLYGVGSSIASWAIKWEYVVLLFLLLADARVCSCLWMMLLISQAEAALEN
LVILNAASLAGTHGLVSFLVFFCFAWYLKGKWVPGAVYTFYGMWPLLLLL-800
```

(N)
LALPQRAYALDTEVAASCGGVVLVGLMALTLSPYYKRYISWCLWWLQYFL
TRVEAQLHVWIPPLNVRGGRDAVILLMCAVHPTLVFDITKLLLAVFGPLW-900
ILQASLLKVPYFVRVQGLLRFCALARKMIGGHYVQMVIIKLGALTGTYVY
NHLTPLRDWAHNGLRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGL-1000
PVSARRGREILLGPADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGR
DKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQM-1100

                        S   T
YTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRG
SLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVEN-1200
LETTMRSPVFTDNSSPPVVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYK

                    L
VLVLNPSVAATLGFGAYMSKAHGIDPNIRTGVRTITTGSPITYSTYGKFL-1300

ADGGCSGGAYDIIICDECHSTDATSILGIGTVLDQAETAGARLVVLATAT
PPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLEVIKGGRHLIFCHSKKKC-1400
DELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGYTGDFDS

    Y   (S)
VIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTQRRGRTGRGKPGIYR-1500
FVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNTPGLPV
CQDHLEFWEGVFTGLTHIDAHFLSQTKQSGENLPYLVAYQATVCARAQAP-1600
PPSWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCMS
ADLEVVTSTWVLVGGVLAALAAYCLSTGCVVIVGRVVLSGKPAIIPDREV-1700
LYREFDEMEECSQHLPYIEQGMMLAEQFKQKALGLLQTASRQAEVIAPAV
QTNWQKLETFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAAVTSP-1800
LTTSQTLLFNILGGWVAAQLAAPGAATAFVGAGLAGAAIGSVGLGKVLID

                    (G)
ILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAILSPGALVVGVVCAA-1900

                                        (HC)
ILRRHVGPGEGAVQWMNRLIAFASRGNHVSPTHYVPESDAAARVTAILSS

FIG. 1-2

SUBSTITUTE SHEET

LTVTQLLRRLHQWISSECTTPCSGSWLRDIWDWICEVLSDFKTWLKAKLM-2000

　　　　　　　　　　　(V)
PQLPGIPFVSCQRGYKGVWRGDGIMHTRCHCGAEITGHVKNGTMRIVGPR
TCRNMWSGTFPINAYTTGPCTPLPAPNYTFALWRVSAEEYVEIRQVGDFH-2100
YVTGMTTDNLKCPCQVPSPEFFTELDGVRLHRFAPPCKPLLREEVSFRVG
LHEYPVGSQLPCEPEPDVAVLTSMLTDPSHITAEAAGRRLARGSPPSVAS-2200
SSASQLSAPSLKATCTANHDSPDAELIEANLLWRQEMGGNITRVESENKV
VILDSFDPLVAEEDEREISVPAEILRKSRRFAQALPVWARPDYNPPLVET-2300

　　　　　　　　　　　　　　　　　　　　　　(S)
WKKPDYEPPVVHGCPLPPPKSPPVPPPRKKRTVVLTESTLSTALAELATR

　　　　　　　　　　　　　　　(FA)
SFGSSSTSGITGDNTTTSSEPAPSGCPPDSDAESYSSMPPLEGEPGDPDL-2400
SDGSWSTVSSEANAEDVVCCSMSYSWTGALVTPCAAEEQKLPINALSNSL
LRHHNLVYSTTSRSACQRQKKVTFDRLQVLDSHYQDVLKEVKAAASKVKA-2500

(F)
NLLSVEEACSLTPPHSAKSKFGYGAKDVRCHARKAVTHINSVWKDLLEDN
VTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVT-2600
KLPLAVMGSSYGFQYSPGQRVEFLVQAWKSKKTPMGFSYDTRCFDSTVTE

　　　　　　　　　　　　　　　　　　　(G)
SDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCGYRRCR-2700
ASGVLTTSCGNTLTCYIKARAACRAAGLQDCTMLVCGDDLVVICESAGVQ
EDAASLRAFTEAMTRYSAPPGDPPQPEYDLELITSCSSNVSVAHDGAGKR-2800
VYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMFAPTLWARMILMTHFF
SVLIARDQLEQALDCEIYGACYSIEPLDLPPIIQRLHGLSAFSLHSYSPG-2900

　　　　　　　　　　　　G
EINRVAACLRKLGVPPLRAWRHRARSVRARLLARGGRAAICGKYLFNWAV

　　　　　　　　　　　　　　　　　　　　　(P)
RTKLKLTPIAAAGQLDLSGWFTAGYSGGDIYHSVSHARPRWIWFCLLLLA-3000
AGVGIYLLPNRQ-3011
　　　　　　　|
　　　Stop codon

　　　　( ) = Heterogeneity due possibly
　　　　　　to 5' or 3' terminal cloning
　　　　　　artefact.
　　　　　　　FIG. 1-3

SUBSTITUTE SHEET

FIG.2—1

5'terminus————
CACTCCACCACCATGAATCACTCCCCTGTGAGGAACTACTGTCTTCACGCAGAAAGGCGTCTAG
CCATGGCGTTAGTATGAGTGTCGTGCAGCCTCCAGGACCCCTCCCGGGAGAGCCATA
GTGGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGAGCACGGGTCCTTCTTGGA
TCAACCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCGCAAGACTGCTAGCCGAGTAGT
GTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAG—300

(Putative initiator methionine codon)
                  G   C
GTCTCGTAGACCGTGCACCATGAGCACGAATCCTAAACCTCAAAAAAACAAACGTAA
CACCAACCGTCGCCCACAGGACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGT
TTACTTGTTGCCGCGCAGGGCCCTAGATTGGGTGTGCGCGACGAGAAAGACTTCCGA
GCGGTCGCAACCTCGAGGTAGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCGAGGGCAG
GACCTGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAATGAGGGCTGCGGGGTGGGC—600
GGGATGGCTCCTGTCTCCCGTGGCTCTCCGGCTAGCTGGGGCCCCACAGACCCCGGCG
TAGGTCGCAATTTGGGTAAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCAT
GGGGTACATACCGCTCGTCGGCGCCCCCTCTTGGAGGCGCTGCCAGGCCCCTGCGCGCATGG
CGTCCGGGTTCTGGAAGACGGCGGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTT

C
CTCTATCTTCCTTCTCGGCCCTGCTCTTGCTTGACTGTGCCCCGGCCTTCGGCCTACCAAGT—900
GCGCAACTCCACGGGGCTTACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTA
CGAGGCGGCGGATGCCATCCTGCCACACTCCGGGTGCGTCCCTGCGTTCGTGAGGGCAA
CGCCTCGAGGTGTTGGGTGGCGATGACCCCTACGTGCTTGTCGGGATGGCAAACTCCC
CGCGACGCAGCTTCGACGTCACATCGATCGGGTCTTCTTTCTTGTCGGGGCCACCCTCTGTTCGGC
CCTCTACGTGGGGGACCTATGCGGGTCTGTTCTTTCGTGCGGCCAACTGTTCACCTTCTC—1200
TCCCAGGCGCCACTGGACGACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAAAC

FIG.2-2

```
                                                              G
GGGTCACCGCATGGCATGGGATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAAT
GGCTCAGCTGCTCCGGATCCCACACGCCATCTTGGACATGATCGCTCGGTGCTCACTGGGG
AGTCCTGGCGGGCCATAGCGTATTTCTCCATGTGGGAACTGGGCGAAGGTCCTGGTAGT
GCTGCTGCTATTTGCCGCGGCGTCGGATTTGTTAGCCTCCTCGACGCGGAAACCCACGTCCTGCGAGGCGGAAGTGCCGGCCA-1500
CACTGTGTCTGGATTTGTTAGCCTCCTCGACGCGGAAACCCACGTCACCGGGGCGGAAGTGCCGGCCA
CAACACCAACGGCAGTTGGCACCTCAATAGCACGGCCCAAGCAGAACGTCCAGCTGAT
CACCGGCTGGGTGGCCAGGGGCTTTTCTATCACCACAAGTTCAACTCTTCAGCTGTCCTGA
GAGGCTAGCCAGCTGCCGACCCCCTTACCGATTTGACCAGGGCGGGCCCTATCAGTTA
TGCCAACGGAAGCGGCCCCGACCAGCGCCCTACTGCTGCACTACCCCCCCAAAACCTTG-1800
CGGTATTGTGCCCGCGAAGAGTGTGTGGTCCGGTATATTGCTTCACTCCCAGCCCCGT
GGTGGTGGAACGACAGGTCGGGCGCGCCCCACCTACAGCTGGGGTGAAAATGATAC
GGACGTCTTCGTCCTTAACAATACCAGGCACCGCTGGGCAATTGGTTCGGTTGTACCTG
GATGAACTCAACTGGATTCACCAAGTGTGGCGAGGCGCTCCTTGTGTCATCGGAGGGGC
GGGCAACAACACCCTGCCCCACTGCTTCCGGACATGCTTCCGCAAGCATCCGGACGCCCACATA-2100
```

```
                                                              C
CTCTCGGTGCGGCTCCGGTCCCCTGGATCACACCCAGGTGCCTGGTCGACTACCCGTATAG
GCTTTGGCATTATCCTTGTACCATCAACTACACCACCATATTAAAATCAGGATGTACGTGGG
AGGGGTCGAACACAGGGGACAGTCCGAAGCTGCCTGCAACTGGACGCGGGGGCGAACGTTGCGATCT
GGAAGACAGGGACAGAGTCCGAGCGTCAGCCGTTACTGCTGACCCGTCCAGCTACACAGTGGCAGGT
CCTCCCGTGTCCTTCACAACCCTACCAGCCTTGTCCAGCCTTGTCCACCGGCCCATCCACCTCCACCA-2400
GAACATTGTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGGGCCAT
TAAGTGGGAGTACGTCGTCTTCCTTCCTTGTTCCTTGCTTGCAGACGCGGCGTCTCCTG
CTTGTGATGATGCTACTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACT
TAATGCAGATGATGCATCCCGGACGCACGTCTTGTATCCTTCCTTGTGTTCTTCTGCTT
TGCATGGTATTTGAAGGGTAAGTGGGTGCCCGGAGGCGGTTACACCTTCTACGGGATGTG-2700
GCCTCCTCCTGTTGGCCGTTGTTGGGGGGTTGTTCGTCGGGTTGATGGCGACACGGAGGT
GGCCGCGTCGTGTGGACGTGGGTTGTCGTCTGCGCTGGTTCTTCGGCTGGTTGGTGGTGGT
TTACAAGGCCTATATCAGCTGGTGGCTGGCTTCAGTATTTTCTGACCAGAGTGGA
```

FIG.2-3

AGGCCAACTGCACGGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGCGGCGAGCGCGTCAT
CTTACTCATGTGTGCTGTACACCGACTCTGGTATTGACATCACCAAATTGCTGCTGGC—3000
CGTCTTCGGACCCCTTTGGATTCTTCAAGCCAGTTGCTTAAAGTACCCTACTTTGTGCG
CGTCCAAGGCCTTCTTCCGGTTCTGCGCGGTTAGCGCGGAAGATGATCGGAGGCCATTACGT
GCAAATGGTCATCATTAAGTTAGGGGCGCTTACTGGCACCTAYGTTTATAACCATCTCAC
TCCTCTTCGGGACTGGGGCGCACAACGGCTTGCGAGATCTGGCCGTGGCTGTGAGCCAGT
CGTCTTCTCCCAAATGGAGACCAAGCTCATCACGTGGGGGCGCAGATACCGCCGGTGCGG—3300
TGACATCATCAACGGCTTGCCTGTTCCGCCGCAGGGGCCGGGAGATACTGCTCGGGCC
AGCCGATGGAATGGTCTCCAAGGGTGGGAGGTTGCTGCGGCCCATCACGGCGTACGCCCA
GCAGACAAGGGCGCCTCCTAGGTCGCATAATCACCAGCCTAACTGTCCCGGGACAAAACCA
AGTGGAGGGTGAGGTCCCAGATTGTGTCAACTGCTCGCCCAAACCTTCCTGGCAACGTGCAT
CAATGGGGTGTGCTGGACTGTCTCTACCACGGCCGGAACGAGGAGGACCATCGCCGTCACCCAA—3600
                                                                        T

GGGTCCTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCGGCCCGCTCC

GCAAGGGTAGCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACCCTTTACCTGGTCAC
GAGGCACGCCGATGTCATTCCCGTGCGCGGGGTGATAGCAGGGGCAGCCTGCTGTC
GCCCCGGCCCATTCCTACTTCCTACTGAAAGGCTCCTCGGGGGGTCCGGGGGTCCGCGGG
GCACGCCGTGGGCATATTTAGGGCCGCGGTGTGCACCCGGTAGTGGGCTAAGGCGGTGGA—3900
CTTTATCCCTGTGGAGAACCTAGAGACAACCATGAGTCCCCGTGTTCACGGATAACTC
CTCTCCACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAG
CGGCAAAAGCACCAAGGTCCCGCTGCAATATGCAGCTCAGGCTATAAGGTGCTAGTACT
CAACCCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGAT
                                                                        T

CGATCCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTC—4200
CACCTACGGCAAGTTCCTTGCCGACGGCGGTGCTCGGGGGGCGCTTATGACATAATAAT
TTGTGACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGA
CCAAGCAGAGACTGCGGGGGCGGAGACTGGTTGTGCTCGCCCACCCCCTCCGGGCTC
CGTCACTGTGCCCCATCCCAACATCCGAGGAGTTGCTCGTCCACCACCGGAGAGATCCC
TTTTACGGCAAGGCTATCCCCTCGAGTAATCAAGGGGGAGACATCTCATCTTCTG—4500
TCATTCAAGAAGAAGTGCGACGAACTGCCGCAAAGCTGTGGGCATCAATGC
CGTGGCCTACTACCGGCGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGT

FIG.2-4

```
                                                              A
CGTGGCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTG
CAATACGTGTGTCACCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGAC
AATCACGCTCCCCCAGGATGCTGTCTCCCGCACTCAACGTGGGGCAGGACTGGCAGGGG-4800
GAAGCCAGGCATCTACAGATTGTGGCACCGGGGAGCGCCCTCCGGCATGTTCGACTC
GTCCGTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGA
GACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGCTTCCGTGCCCAGGACCA
TCTTGAATTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATC
CCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGTAGCGTACCAAGCCACCGTGTG-5100
CGCTAGGGCTCAAGCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCT
CAAGCCCACCCTCCATGGCCAACACCCCGCTATACAGACTGGGCGCTGTTCAGAATGA
AATCACCCTGACGCACCAGTCACCAAATACATCATGACATGTCGGCCGACCTGGA
GGTCGTCACGAGCACAGGCTGCGTCGTTGGCGGCGTCCTGGCGCTCTTGCGCCGCGTATTG
CCTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAAT-5400
CATACCTGACACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGAGAAGAGTGCTCTCAGCA
CTTACCGTACATGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGG
CCTCCTGCAGACCGGTCCCGTCAGGCAGAGGTTATCGCCCCTGTCTCCAGACCAACTG
GCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATA
CTTGGCGGGCTTGTCAACGCTGGCCTGGTAACCCGCCATTGCTTCATTGATGGCTTTTAC-5700
AGCTGCTCGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGG
GTGGGTGGCTGCCCAGCTGCGCCAGCTCGCGCCCCGCTTGTGGGCGCTGGCTT
AGCTGGCGCCGCGCCATCGCGCAGTGTTGGACTGGGGCGGAAGGTCCTCATAGACATCCTTGCAGG
GTATGGCGGCGGCTGGGCGGAGCTCTTGTGCATTCAAGATCATGAGCGGTGAGGTCCC
CTCCACGGAGGACCTGGTCAATCTACTGCGCCGGCCATCCTCCGCCCGGAGCCCTGTAGT-6000
CGGCGTGGTCTGTGTGCAGCAGCAATACTGCGCGCCGCGGGCGCGAGGGGGCAGTGCA
GTGGATGAACCGGCTGATAGCCTTCGCCCGGCTCACTGCCATGTTCCCCACGCACTA
CGTGCCGGAGAGCGATGCGACTGCACCAGTGGATAAGCTCGAGTGTACCACTCACTGTAAC
CCAGTCCTGAGGCGACTGCACCACTGGGACTGTTGAGCGACTTTAAGACCTG-6300
TTCCTGGCTAAGGACATCTGGGACATGCCCTGGGATCCCCCTTTGTCCTGCCAGCGCGG
GCTAAAAGCTAAGCTCATGCCCACAGCTGCCACAGCTGCCACAGCTCTGTCCTGCCAGCGCGG
```

FIG.2-5

```
GTATAAGGGGGTCTCGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGA
GATCACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAA
CATGTGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGCCCCTGTACCCCCTTCC
TGCGCCGAACTACACGTTCGCGCTACGTGACGGGTGTCTGCAGAGGAATATGTGAGATAAG-6600
GCAGGTGGGGACTTCCACTCGCCCGAATTTTCACAGAATTGGACGGGGTGCGCCTACAATCTCAAATGCCCGTG
CCAGGTCCCATGCCAAGCCCTTGCTGCGGGAGGAGTATCATTCAGAGTAGGACTCCACGAATA
GCCCCCCTGCCAAGCCCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCAT
CCCGGTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGCCGAAGGTTGGCGAGGGGATC-6900
GCTCACTGATCCCCTCCCATATAACAGCAGAGGCCGGCCGTATCCGCTCCATCTCTCAAGGCAAC
ACCCCCCTCTGTGGCCAGCTCCTCGCTAGCCAGCTATCCGAGCTCATAGAGGCCAACCTCCTATGGAG
TTGCACCGCTAACCATGACTCCCCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTCGGA
GCAGGAGATGGGCGCAACATCACCAGGGGTTGAGTCAGAAAACAAAGTGGTGATTCTCGGA
CTCCTTCGATCCGCTTGTGGCGGAGGAGGACGAGGGATCTCCGTACCGCAGAAAT
CCTGCGGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTGGGCCGGCCGACTATAA-7200
CCCCCCGCTAGTGGAGACGTGGAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTG
TCCGCTTCCACCTCCAAAGTCCCCTCCTGTGCCTGGCCTCGGAAGAAGCGGACGGTGGT
CCTCACTGAATCAACCCTATCTACTGCCTGGCCGACAATCCTCTGAGCCCGCCCCTTC
CTCCTCAACTTCCGGCATTACGGGCGACAATACGACACAATCCTCTGAGCCCGCCCCTTC
TGGCTGCCCCCGACTCCGACGCTGAGTCCTATTCCCATGCCCCCCCTGGAGGGGGA-7500
GCCTGGGGATCCGGGATCTTAGCGACGGGTCAGCGGTCAGTAGTGAGGCCAACGC
GGAGGAGTGTCGTGTGTCGTCAATGTCTTACTCTTGGACAGCGCCACTCGTCACCCCGTG
CGCGCGGAAGAACAGAAACTGCCCATCAATGCCATCAAGCAACTCGTTGCTACGTCACCA
CAATTGGTGTATTCCACCACCTCACGCAGTGCTGCCAAGGCAGAAGAAAGTCACATT
TGACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGC-7800
GGCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCC
ACACTCAGCCAAATCCAAGTTTGGTTATGGGCAAAGACGTCCGTTGCCATGCCAGAAA
GGCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGAAGACAATGTAACACCAAT
AGACACTACCATCATGCCTAAGAACGAGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGTCG
```

FIG.2-6

```
TAAGCCAGCTCGTCTCATCGTGTTCCCGATCTGGGCGTGCGCGTGTGCGAAAGATGGC-8100
TTTGTACGACGTGGTTACAAAGCTCCCCTTGCCCGTGATGGGAAGCTCCTACGGATTCCA
ATACTCACCAGGACAGCAGCGGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAACCCC
AATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCG
TACGGAGGAGGCAATCTACCAATGTTGTGACCCTGACCCTGACCCCGGCGTGGCCATCAA
GTCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTG-8400
CGGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCAC
TTGCTACATCAAGGCCCGGCAGCCTGTCGAGCCCAGGGCTCCAGGACTGCACCATGCT
CGTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGC
GAGCCTGAGAGCCTTCACGGAGGCTATGACCAGTACTCCGCCCCCCCCGGGACCCCCC
ACAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCTCCAACGTGTCAGTCGCCCA-8700
CGACGGGCTGGAAAGAGGGTCTACTACTCACCCGTGACCCTACAACCCCCCTCGCGAG
AGCTGCGTGGGGAGGAGCAAGACAGCACACTCCAGTCAATTCCTGGCTAGGCAACATAATCAT
GTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTAT
AGCCAGGGACCAGCTTGAACAGGCCCTGATTGCGAGATCCAGATCTACGGGGCCTGCTACTCCAT
AGAACCACTTGATCTACCTCCAATCATTCAAAGACTCCAAAGACTCCATGGCCTCAGCGCATTTTCACT-9000
CCACAGTTACTCTCCAGTGAAATTAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGT

G
ACCGCCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTCGGCCAG
AGGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCT
CAAACTCACTCCAATAGCGGCCGCTGGCCAGCTGACTTGTCCGGCTGTGTTCACGGCTGG
CTACAGCGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCGGCGGATCGGTTGGTT-9300
TTGCCTACTCCTGCTGCTCAGGGTAGGCATCTACCTCCTCCCCAACCGATGAGGTT
GGGGTAAACACTCCGGCCT----------------------------3'terminus
```

Some clonal heterogeneities producing amino acid substitutions are shown. There are many other "silent"mutations(not shown).

FIG.3-1   Human 23

```
1    GlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyLysAlaProLeuGlyGlyArgAla
     GGCTTCGCCGACCTCATGGGGTACATACCGTCGGCGTCCGGCGCCCCTCTTGGAGGCCGTGCC

61   ArgAlaLeuAlaHisGlyValArgValLeuGluAspGlyValAsnTyrAlaThrGlyAsn
     AGGGCCCTGGCCACGGCGTCCGGGTTTGGAAGACGGGCTGAACTATGCAACAGGAAC
            CG                                  A

121  LeuProGlyCysSerPheSerIlePheLeuLeuAlaAlaLeuSerCysLeuThrValPro
     CTTCCTGGTTGCTCCTTTCTATCTTCCTTCCTTCCTGGCCCTACTCTTCCTGACCGTGCCC
                                   GA                    T

181  AlaSerAlaThrTyrGlnValArgAsnSerThrGlyLysLeuTyrHisValThrAsnAspCysPro
     GCTTCAGCTACCAAGTGCGCAACTCTACGGGCTTTACCATGTCACCAATGATTGCCCT

241  AsnSerSerIleValTyrGluAlaAlaAlaAspAlaIleLeuHisAlaProGlyCysValPro
     AACTCGAGTATTGTGTACGAGGCGGCCGATGCCATCCTGCACGCTCCGGGGTGTCCCT
                            T                                  C

301  CysValArgGluAspAspAsnValSerArgArgCysTrpValAlaValThrProThrValAlaThr
     TGGCGTTCGCGGAGGATAACGTCTCGAGAGTGTTGGGGTGGCGTGACCCCCACGGTGGCCACC
              G                                               T

361  LysAspGlyLysLeuProThrThrGlnLeuArgArgHisIleAspLeuLeuValGlySer
     AAGGACGGCAAACTCCCACACCCCCAGCTTCGACGTCACATCGATCTGCTTGTCGGGAGC
                                                         A

421  AlaThrLeuCysSerAlaLeuTyrValGlyAspLeuCysGlySerIlePheLeuValGly
     GCCACCCTCTGCTCGGCCCTCTACGTGGGGGACCTTTGCGGGTCCATCTTTCTTGTCGGT
             T

481  GlnLeuPheThrPheSerProArgArgHisTrpThrThrGlnAspCysAsnCysSerIle
     CAACTGTTTACCTTCTCTCCCAGGCGCCACTGGACGACGCAGGACTGCAACTGTTCTATC
                                                         C
```

FIG.3-2

```
      TyrProGlyHisIleThrGlyHisArgMetAlaTrpAspMetMetMetAsnTrpSerPro
541   TATCCCGGCCATATAACGGGTCACCGGCATGGCATGGGATATGATGATGAACTGTCCCCT
                                                              G

      ThrAlaAlaLeuValValAlaAlaGlnLeuLeuArgIleProGlnAlaIleLeuAspMetIle
601   ACGGGGCATTGGTAGTAGCTCAGCTGCTCCGGATCCCACAGCCATCTTGGACATGATC
            G   AG

      AlaGlyAlaHisTrpGlyValValLeuAlaGlyMetAlaTyrPheSerMetValGlyAsnTrp
661   GCTGGTGTCACTGGGGAGTCCTGGCGGGCATGGCGTATTTCTCCATGGTGGGGAACTGG
                                            G

      AlaLysValLeuValValLeuLeuLeuPheAlaGlyValAspAlaGluThrHisArgThr
721   GCGAAGGTCCTGGTAGTGCTTCTATTGCCGGTCGACGCGGAAACCCACCGTACC
                                                        G

      GlyGlySerAlaAlaArgSerThrAlaGlyValAlaSerLeuPheThrProGlyAlaArg
781   GGGGGAAGTGCCCGCCCCAGCAGCGGCTGGAGTTGCTAGTCTCTTCACACCCAGGCGCTAGG
      C       T    A

      GlnAsnIleGlnLeuIleAsnThrAsnGlySerTrpHisIleAsnSerThrAlaLeuAsn
841   CAGAACATCCAGCTGATCAACACGCAGTTGGCACATCAATAGTACGGCCTTGAAC
                                                        AT

      CysAsnAspSerLeuThrThrGlyTrpLeuAlaGlyLeuPheTyrHisHisLysPheAsn
901   TGCAATGACAGCCTTACCACGGGTTAGCGGGGCTTTTCTATCACCATAAATTCAAC
          A                                         A

      SerSerGlyCysProGluArgLeuAlaSerCysArgProLeuThrAspPheAlaGln
961   TCTTCAGGCTGTCCCGAGAGTTGGCCAGCTGCCGACCCCTCACCGATTTGCCCAGG
      G   A                                           G
```

FIG.4—1

Human 27

```
     GlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyGlyAlaAla
  1  GGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGGGGCGCTGCC

     ArgAlaLeuAlaHisGlyValAspGlyValLeuGluAspGlyValAsnTyrAlaThrGlyAsn
 61  AGGGCCCTGGCGCATGGCGTCCGGGGTTCTGGAAGACGGCGTGAACTATGCAACAGGGAAC

     LeuProGlyCysSerPheSerIlePheLeuLeuAlaLeuLeuSerCysLeuThrValPro
121  CTTCCTGGTTGCTCTTTCTCTATTCCTTCTGCTCTCTTCTGCTCCGGTGCCC

     AlaSerAlaTyrGlnValArgAsnSerSerGlyIleTyrHisValThrAsnAspCysPro
181  GCATCGGCCTACCAAGTACGCAACTCCTCGGGCATTTACCATGTCACCAATGATTGCCCT
                                                                  C

     AsnSerSerIleValTyrGluThrAlaAspThrIleLeuHisSerProGlyCysValPro
241  AATTCGAGTATTGTGTACGAGACGGCCGACACCATCTCACTCTCCGGGGTGTGTCCCT
                                                      C

     CysValArgGluGlyAsnAlaSerLysCysTrpValProValAlaProThrValAlaThr
301  TGCGTTCGCGAGGGTAACGCCTCGAAAGTTGGGTGCCGGTAGCCCCACAGTGGCCACC
                                                      G

     ArgAspGlyAsnLeuProAlaThrGlnLeuArgArgHisIleAspLeuLeuValGlySer
361  AGGGACGGCAACCTCCCCGCAACGCCAGCTTCGACGTCACATCGATCTGCTTGTCGGGAGT
                                                      G         G

     AlaThrLeuCysSerAlaLeuTyrValGlyAspLeuCysGlySerValPheLeuValGly
421  GCCACCCTTTGCTCGGCCCTCTATGTGGGGGACTTGTGCGGGTCGTCTTTCTTTGTCGGT
                      C                                         C

     GlnLeuPheThrPheSerProArgArgHisTrpThrThrGlnAspCysAsnCysSerIle
481  CAACTGTTCACTTTCTCCCCGGACGGCCACTGGACAACGCAAGATTGCAACTGCTCTATC
                                                              A
```

FIG.4-2

541 TyrProGlyHisIleThrGlyHisArgMetAlaTrpAspMetMetAsnTrpSerPro
TACCCCGGCCATATAACGGGACACCGGCATGGCATGATGATGAACTGTCCCT

601 ThrAlaAlaLeuValMetAlaGlnLeuLeuArgIleProGlnAlaIleLeuAspMetIle
ACAGCAGGCTGGTAATGGCTCAGCTGCTCAGGATCCCGCAAGCCATCTTGGACATGATC
G

661 AlaGlyAlaHisTrpGlyValLeuAlaGlyIleAlaAlaTyrPheSerMetValGlyAsnTrp
GCTGGTGCTCACTGGGGAGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGG

721 AlaLysValValValLeuLeuPheAlaGlyValAspAlaThrThrTyrThrThr
GCGAAGGTCCTGGTGGTGGTCGTTTGCGTGTTTGCCGGCGTCGATGCGACAACCTATACCACC

781 GlyGlyAsnAlaAlaArgThrThrGlnAlaLeuThrSerPhePheSerProGlyAlaLys
GGGGGAAATGCTGCCAGGACCACGGCAGGCGCTCCACCAGTTTTTCAGCCCAGGCGCCAAG

841 GlnAspIleGlnLeuIleAsnThrAsnGlySerTrpHisIleAsnArgThrAlaLeuAsn
CAGGATATCCAGCTGATCAACACGGCAGTTGGCACATCAATCGCACGGCCTTGAAC
G   T

901 CysAsnAlaSerLeuAspThrGlyTrpValAlaGlyLeuPheTyrTyrHisLysPheAsn
TGTAATGCCAGCCTCGACACTGGCTGGGTAGCCGGGCTCTTCTATTACCACAAATTCAAC
T   G

961 SerSerGlyCysProGluArgMetAlaSerCysArgProLeuAlaAspPheAspGln
TCTTCAGGCTGCCCGGAGAGGATGGCCAGCTGTAGGCCCCTTGCCGATTTCGACCAGG
C

SUBSTITUTE SHEET

FIG.5-1

1. human 27   2. HCV 1   3. human 23

1 CGGCTTCGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCtCCTCTTGGGGCCTGCCAGGCCCTGGC
********* *** ************* * ******* ****** *** *** ** ***** **** ***
1 CGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCCCTCTTGGAGGCCTGCCAGGCCCTGGC
********* *** ************* * ******* ***** *** *** ** ***** *** * ***
1 CGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGCGCCCCTCTTGGAGGCcgTGCCAGGCCCTGGC

73 GCATGGCGTCCGGGTTCTGGAAGAGACGGGCCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTAT
***** *** ************* ************* ******* ************* ******** *** *
73 GCATGGCGTCCGGGTTCTGGAAGAGACGGGCCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTAT
***** *** ************* ************* ******* ***** ******* ** * *** ***
73 GCAcGGCGTCCGGGTTtTGGAAGAGACGGGCCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCtCCTTtTCTAT

145 CTTCCTTCTGGCTCTCTCTTGCCtGACCGGTGCCCGCAtCGGCCTACCAAGTACGCAACTCCtCGGGcaT
*** ******* **** *** *** **** ** ** ** ****** **** * *
145 CTTCCTTCTGGCCCCTGCTCTCTTGCCtGACtGTGCCCGCGCTTCGGCCTACCAAGTGCGCAACTCCACGGGGCT
*** ******* *** *** ******* ** *** ******* ******* ******* ** * ***
145 CTTCCTTCTGGCCCTaCTCTCTTGCCtGACCGGTGCCCGCTTCagCCGTGCCCACCAAGTGCGCAACTCtACGGGGCT

217 TTACCAtGTCACCAATGATTGCCCTAAtTCGAGTATTGTGTACGAGACGGGCGAcaCCATCCTaCACtCTCC
***** * *** ***** *** **** ***** ** ** * **** *** ***
217 TTACCAcGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATCCTGCACaCTCC
***** * *** ***** *** **** **** ******* ** *** ***** ******* ******
217 TTACCAtGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATCCTGCACGCTCC

289 GGGGTGtGTCCCTTGCGTTCGtGCGGAGGGtAACGCCTCGAaaTGTTGGGTGcCGgTagCCCCaCACAGTGGCCAC
***** * ** ***** ** ** ****** **** ***** * ** * ****** *** ***
289 GGGGTGcGTCCCTTGCGTTCGtGCGGAGGGcAACGCCTCGAGGTGTGGGTGGCGaTGACCCCtACGGTGGCCAC
***** * ** ***** ** ** ****** * ***** *** ****** * ******** ***
289 GGGGTGtGTCCCTTGCGTTCGcGCGGAGGatAACGCtCTCGAGGatAACGTGTTGGGTGGCGaGtGACCCCcACGGTGGCCAC

FIG.5-2

```
361  CAGGGAcGGCAAcCTCCCGCaACGCAGCTTCGACGTCACATCGATCTGCTTGTGTCGGGAGtGCCACCCTtTG
     **  *****  *   ****** ****************************************  **** ** **

361  CAGGGAtGGCAAAACTCCCGcGACGCAGCTTCGACGTCACATCGATCTGCTTGTCGGGAGCGCCACCCTCTG
     **  *  ****  *** ****  ******************************************** *** ***

361  CAaGGAcGGCAAACTCCCaCaACaACGCAGCTTCGACGTCACATCGATCTGCTTGTCGGGAGCGCCACCCTCTG
     **  *  ****  ***  * *  *************************************  **  ** ** ***

433  CTCGGCCCTCTAtGTGGGGACTtGTGCGGGTCTCGTCTTTCTtGTCGGtCAACTGTTCACtTTCTCCCCAG
     *********  *  ************  ***********  ** ***** ******** **  ** *** **

433  tTCGGCCCTCTACGTGGGGACCTGTGCGGGTCTCGTCTtTCTGTCGGcCAACTGTTCACCTTCTCTCCCAG
     *********** ***********  ***********  ** *** ***** ******** ** ***** **

433  cTCGGCCCTCTACGTGGGGACCTtGCGGGTCCaTCTTTCTtGTCGGtCAACTGTTtACTTTCTCTCCCAG
     *********** ***********  *****  * *****  ** ***** ******  ** ****** **

505  GCGCCACTGGACAaCGCAAGAtTGCAACTGCTCTATCTAcCCGGCCCATATAACGGaCACCGCATGGCATG
     *********** * ***** * ***** *** ****** *  ************** * ************ **

505  GCGCCACTGGACGACGCAAGgTTGCAAtGCTCTATCTATCCGGCCCATATAACGGTCACCGCATGGCATG
     ***********  ***** *  **** * ************************************** ** ***

505  GCGCCACTGGACGACGCAGGaCtGCAAcTGtTCTATCTATCCGGCCCATATAACGGTCACCGCATGGCATG
     ***********  ***** *  **** * * ************************************** ** ***

577  GGATATGATGATGAACTGGTCCCCTACAgCaGCGCtGGTAAtGGCTCAGCTGCTCAGGATCCCgCAAGCCAT
     **********  ***********  * * **** * * *** **  *** *********** *   *******

577  GGATATGAtGATGAACTGGTCCCCTACGaCGGCGGTTGGTAAtGGCTCAGCTGCTCCGGATCCCCACAAGCCAT
     **********  ***********  *  ***** *  *** **  ******* *** ********* ********

577  GGATATGAtGATGAACTGGTCCCCTACGcGGCGGCaTTGGTAgTAgCTCAGCTGCTCCGGATCCCCACAAGCCAT
     **********  ***********  *  ***** *  *** **  ******* *** ******************

649  CTTGGACATGATCGCTGGTGCTGTGCTCACTGGGGAGTCCTaGCGGGCATAGCGTATTTCTCCATGGTGGGAACTG
     ********************** *  ************** * ********** ******** *  ************

649  CTTGGACATGATCGCTGGTGCTGTGCTCACTGGGGAGTCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGAACTG
     *********************  * ***** ************ ********** ******** ** ***********

649  CTTGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATgGCGTATTTCTCCATGGTGGGAACTG
     ********************** *  ************ ****** * ****** ****** ** ***********

721  GGCGAAGGTCCTGGTgGTGCTGtGCTgTTGCCGGCGTCGAtGCGacAACCtAtacCACCGGGGgAaTGC
     ************** * ****** *** * ********** * ** * ** *  *  ** ***** * * ***

721  GGCGAAGGTCCTGGTAGTGCTGtGCTATTTGCCGGCGTCGACGCGAAACCCACgtCACCGGGGAAGTGC
     *************** ****** *** *** ************ ***** ** ** ****** ***** ***

721  GGCGAAGGTCCTGGTAGTGCTGttCTATTTGCCGGCGTCGACGCGGAAACCCACCgtACCGGGGGAAGTGC
     *************** ****** *  *** *** ************ **** ** ** *  ** ***** ***
```

```
793  tGcCaggaACcacGcaggGcgcTcaccAGtTtTtCagcGCGCCAAGGCGCCAGGAtaTCCAGCTGATCAACAC
          *        *       *   *         *  *      **  *    ***********
793  CGgCCaCaCTgtGtCTGGAtTTGTtAGcCCTCcTCgCACCAGGCGCCAAGCAGAACgTCCAGCTGATCAACAC
     ** ** **  *   *  *   *     **** ***   *  *      *    *    *************
793  CGcCCgCAgcacGgCTGGAgTTGcTAGtCTCtTCacCACCAGGCGCtAgGCAGAACaTCCAGCTGATCAACAC

865  CAACGGCCAGTTGGCACaCaTCAAtCgCACGGCCtTGAACTGtAATGcgAGCCTCgACACTgGGCTGGGTaGCGgg
     *********************** ** ******  **   **     *  ***  *    ** *  * * **
865  CAACGGCCAGTTGGCACCtCAATAGCACGGCCCTGAACTGCAATGAtAGCCTCAACACCGGCTGGTTgGCaGG
     ********************    ******  *     ** *   ** *  ***  ** *  * **  **
865  CAACGGCCAGTTGGCACaTCAATAGtACGGCCtTaCCACCGGCTGGTTaGCggg

937  GCTcTTCTATtACCACAaaTTCAAcTCTTCAGGCTGcCCCGAGaGGAtGCCAGCTGtaGcCCCTTgCCGA
     ***             *  *************************  **    * *   * ***  ***
937  GCTTTTCTATcACCACAAgTTCAAcTCTTCAGGCTGtCCtGAGAGGcTaGCCTGCCAGCCCCCTTACCGA
     ***   *  **  **  * *************************  *       *    * ***  ***
937  GCTTTTCTATcACCAtAAaTTCAAcTCTTCAGGCTGtCCCGtGtCCCGAGAGGTtGGCCAGCTGCCGACCCCCTCACCGA

1009 TTTcGACCAGG
     ***  ******
1009 TTTTGACCAGG
     ****  *****
1009 TTTTGcCCAGG
```

FIG.5-3

FIG.6

```
1   GFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSsGi
       *********************************************************************  *
1   GFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSTGL
       *****************  **************************************************** *
1   GFADLMGYIPLVGAPLGGrARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSTGL

73  YHVTNDCPNSSIVYEtAdtILHsPGCVPCVREGNAskCWVpvaPTVATRDGnLPATQLRRHIDLLVGSATLC
       ***************  ** *** **********   ***   ****** **  ********************
73  YHVTNDCPNSSIVYEAADAILHtPGCVPCVREGNASrCWVAmTPTVATRDGKLPATQLRRHIDLLVGSATLC
       ****************  ****** **  ****** * ****  ***** ** ********************
73  YHVTNDCPNSSIVYEAADAILHaPGCVPCVREdNvSRCWVAvTPTVATkDGKLPtTQLRRHIDLLVGSATLC

145 SALYVGDLCGSVFLVGQLFTFSPRRHWTTQdCNCSIYPGHITGHRMAWDMMMNWSPTtALVMAQLLRIPQAI
       ********** ***************** * **********************************  ***********
145 SALYVGDLCGSVFLVGQLFTFSPRRHWTTQgCNCSIYPGHITGHRMAWDMMMNWSPTtALVMAQLLRIPQAI
       ********** ***************** * **********************************  *** ******
145 SALYVGDLCGSiFLVGQLFTFSPRRHWTTQdCNCSIYPGHITGHRMAWDMMMNWSPTaALVvAQLLRIPQAI

217 LDMIAGAHWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDAtTyTGGnAarTtqaltSffsPGAKQdiQLINT
       ************************************   *  ***          *     ***** ** *****
217 LDMIAGAHWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDAEThVTGGSAghTvsGfvSLlaPGAKQNvQLINT
       ************** ************************** *  ****   *   *   *    **** *****
217 LDMIAGAHWGVLAGmAYFSMVGNWAKVLVVLLLFAGVDAEThrTGGSAarstaGvaSLftPGArQNiQLINT

289 NGSWHiNrTALNCNaSLdTGWvAGLFYyHKFNSSGCPERmASCRPLaDFDQ
       ***** *  ***** ** *** * *** **************** *** *** *
289 NGSWHiNSTALNCNDSLnTGWLAGLFYHHKFNSSGCPERLASCRPLTDFDQ
       ***** ***************** **************************** *
289 NGSWHiNSTALNCNDSLtTGWLAGLFYHHKFNSSGCPERLASCRPLTDFaQ
```

1. human 27
2. HCV 1
3. human 23

CLUSTERED PAIR-WISE 'REGION' ALIGNMENT
in 'identity (no translation)' alphabet of:

1. ssThorn#8.r (1-587)
2. ssEC1#2.r (1-587)
3. ssHCT18#7.r (1-587)
4. env1.hcv (1-1657)

FIG.7-1

```
1
1
1
1

289  gggtgggcgggatggctcctgtctcccgtggctctcggcctagctgggcccacagacccccggcgtaGg

  3  ATTCGCAATTTGGGTAAGGTCATCGATACCCTTACGTGCGGGCTTCGCCGACCCTCATGGGGTACATACCGCTC        GA
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||        ||
  3  ATTCGCAATTTGGGTAAGGTCATCGATACCCTTACGTGCGGGCTTCGCCGACCCTCATGGGGTACATACCGCTC        GA
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||        ||
  3  ATTCGCAATTTGGGTAAGGTCATCGATACCCTTACGTGCGGGCTTCGCCGACCCTCATGGGGTAtATACCGCTC        GA
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||        |
361  tcgCGCAATTTGGGTAAGGTCATCGATACCCTTACGTGCGGGCTTCGCCGACCCTCATGGGGTACATACCGCTC

 75  GTCGGGCGCCCCTCTTGGGgGGGCGCTGCCCTGGCGCTGCCCAGGGCCCTGGCGCTCCGGGTTCTGGAAGACGGCGTGAAC
     ||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 75  GTCGGGCGCCCCTCTTGGAGGGGCGCTGCCCTGGCGCTGCCCAGGGCCCTGGCGCTCCGGGTTCTGGAAGACGGCGTGAAC
     ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 75  GTCGGGCGCCCCTCTTGGAGGGGCGCTGCCCTGGCGCTGCCCAGGGCCCTGGCGCTCCGGGTTCTGGAAGACGGCGTGAAC
     ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
433  GTCGGGCGCCCCTCTTGGAGGGGCGCTGCCCTGGCGCTGCCCAGGGCCCTGGCGCTCCGGGTTCTGGAAGACGGCGTGAAC
```

FIG.7-2

```
147  TATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTCTTCCTTCGGCCCTGCTCTCTTGtCTGACcGTG
     ||||||||||||||||||||||||||||| | ||  ||||||||||||||||||||||||| ||||||| 
147  TATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTtTATCTTCCTTCGGCCCTGCTCTCTTGCTTGACTGTG
     |||||                      | |||||  |||||||||||||||||||||||||||||||||||
147  TATGC  CAGGGAACCTTCCTGGTTGCTCTTTCTTTCTCTATCTTCCTTCGGCCCTGCTCTCTTGCCTGACTGTG
     |||||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
505  TATGCAACAGGGAACCTTCCTGGTTGCTCTTTCCTGGTCTCTTTCTCTATCTTCCTTCGGCCCTGCTCTCTTGCTTGACTGTG

219  CCCGCTTCAGCCTACCAAGTGCGCAACTCCaCGGGGCTTTACCATGTCACCAACGATTGCCCAACTCGAGt
     ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||| 
219  CCCGCTTCAGCCTACCAAGTGCGCAACTCCtCGGGGCTTTACCATGTCACCAATGATTGCCCtAACTCGAGc
     |||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||| 
219  CCCGCTTCAGCCCACCAAGTGCGCAACTCCACGGGGCTTTACCATGTCACCAATGATTGCCCCAACTCGAGT
     |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
577  CCCGCTTCgGCCtACCAAGTGCGCAACTCCACGGGGCTTTACCACGTCACCAATGATTGCCCtAACTCGAGT

291  ATTGTGTACGAGGGCGGCCGATGCtATCCTGCACgCTCCGGGGGTGTCCCTTGCGTTCgCGAGGGtAACGCC
     ||||||||||||||||||||||||| |||||||| |||||||||||||||||||||||||||| ||||||||
291  ATTGTGTACGAGGGCGGCCGATGCCCCATCCTGCACACTCCGGGGGTGTCCCTTGCGTTCCCTCGCACGAGGGCAAACGTC
     |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
291  ATTGTaTACGAaGCGGCCGACGCCATCCTGCACACTCCGGGGGTGTCCCTTGCGTTCCCTTGCACGAGGGCAAACGTC
     |||| ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
649  ATTGTgTACGAGagGCGGCCGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCCGTTCgTGAGGGCAAACGCC

363  TCGAGGTGTTGGGTGGGCGATGACCCCCACGGTGGCCGCCAGGGaCGGCCAgACTCCCACAACGCAGCTgCGA
     ||||||||||||||||||||||||||||||||||||||| ||||||||| |||||||||||||||||||||||
363  TCGAGGTGTTGGGTGGGCGATGACCCCCACGGTGGCCGCCAGGGgCGGCCAAACTCCCACAACTCCCACGCAGCTTCGA
     ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
363  TCGAGGTGTTGGGTGGGCGGtGACCCCCACGGTGGCCGCCAGGGaTGGCCAAACTCCCACAACGCAGCTTCGA
     ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
721  TCGAGGTGTTGGGTGGGCGGaTGACCCCtACGGTGGCCCCACCAGGGATGGCAAACTCCCgACGCAGCTTCGA
```

435 CGTCACATCGATCTGCTTGTCGGGAGCGCcACCCCTCTGCTCGGCCCCTCTACGTGGGGGACCTGTGCGGGTCc
|||||||||||||||||||||||||||| |||||||||||| |||||||||||||||||||||| ||||||||||
435 CGTCACATCGATCTGCTTGTCGGGAGCGCtACCCCTCTGCTCGGCCCCTCTACGTGGGGGACCTGTGCGGGTCT
|||||||||||||||||||||||||||| |||||||||||||||||||||| |||||| ||||||||||||||
435 CGTCACATCGATCTGCTTGTCGGGAGCGCCACCCCTCTGCTCGGCCCCTCTAtGTGGGGGACtTGTGCGGGTCT
|||||||||||||||||||||||||||| ||| ||||||||| |||||||| |||||||||| |||||||||
793 CGTCACATCGATCTGCTTGTCGGGAGCGCCACCCCTCTGtTCGGCCCCTCTACGTGGGGGACcTaTGCGGGTCT

507 aTCTTtCTTGTGTCGGTCAACTGTTcACCTTCTCTCCCAGGCGCCACTGGACGACGCGCAAGGTTGCAATTGCTCT
||||| ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
507 GTCTTcCTTGTGGTCAACTGTTcAACCTTCTCTCCCAGGCGCCACTGGACGACGCGCAAGGTTGCAATTGCTCT
||||| ||||||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
507 GTCTTTCTTGTGGGCCAACTGTTTACCTTCTCTCCCAGGCGCCACTGGACGACGCGCAAGGTTGCAATTGCTCT
||||| ||||||| ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
865 GTCTTTCTTGTGTGGGCCAACTGTTcACCTTCTCTCCCAGGCGCCACTGGACGACGCGCAAGGTTGCAATTGCTCT

579 ATCGAATTC
||||||||
579 ATCGAATTC
|||||||
579 ATCGAATTC
||| |
937 ATCtAtccC

SUBSTITUTE SHEET

FIG.7-3

```
                       10        20        30        40
EC10          GAATTCGGACGACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATAT
              X::::::::::::::::::::::::::::::::::::::::::::::::::
HCV1   CTCTCCCAGGCGCCACTGGACGACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATAT
              550       560       570       580       590       600

        50        60        70        80  A     90       100
       AACAGGTCACCGCATGGCATGGGATATGATGATGAACTGGTCCCCTACGACGGCGTTAGT
       ::: ::::::::::::::::::::::::::::::::::::::::::::::::::::: ::
       AACGGGTCACCGCATGGCATGGGATATGATGATGAACTGGTCCCCTACGACGGCGTTGGT
          610       620       630       640       650       660

       110       120       130       140       150       160
       GGTAGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTG
        : :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
       AATGGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTG
          670       680       690       700       710       720

       170       180       190       200       210       220
       GGGAGTCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCTTGGC
       :::::::::::::::::::::::::::::::::::::::::::::::::::::::::  :::
       GGGAGTCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGT
          730       740       750       760       770       780

       230       240       250       260       270       280
       AGTGCTGCTGCTATTTGCCGGCGTCGACGCGGAAACCCACGTCACTGGGGGGATCGCCGC
       :::::::::::::::::::::::::::::::::::::::::::::::::::::  ::: :   ::::
       AGTGCTGCTGCTATTTGCCGGCGTCGACGCGGAAACCCACGTCACCGGGGGGAAGTGCCGG
          790       800       810       820       830       840

       290       300       310       320       330       340
       CAAAACTACGGCTAGCCTTACTGGTCTCTTCAATTTAGGTGCCAAGCAGAACATCCAGCT
       : : ::::   : ::   ::  :  : ::  ::    :::  ::::::::::    :::::::::
       CCACACTGTGTCTGGATTTGTTAGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAGCT
          850       860       870       880       890       900

       350       360       370       380       390       400
       GATCAACACCAACGGCAGTTGGCACATCAACAGGACGGCCTTGAACTGCAATGATAGCCT
       :::::::::::::::::::::::::::::::  ::::  ::  :  :::::::  ::::::::::::::::::::
       GATCAACACCAACGGCAGTTGGCACCTCAATAGCACGGCCCTGAACTGCAATGATAGCCT
          910       920       930       940       950       960

       410       420
       CAACACCGGCTGGAATTC
       :::::::::::::X
       CAACACCGGCTGGTTGGCAGGGCTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCC
          970       980       990      1000      1010      1020
```

FIG.8

AA #117-308 (putative envelope region) FIG.9

1) HCT #18 (USA)   3 clones sequenced
2) JH23 (USA)   ?
3) JH 27 (USA)   ?
4) PBL-Th (USA)   2 clones sequenced
5) EC1 (Italy)   3 clones sequenced
6) HCV-1 (chimpanzee)   multiple

```
    C/M←─┬─→S
1)                                    (P)
2)
3)
4)
5)
6)RNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNL


1)                        H
2)
3)                                S                    T   T
4)         L
5)    (F)                          S
6)PGCSFSIFLLALLSCLTVPASAYQVRNSTGLYHVTNDCPNSSIVYEAADAILH


1)       Y
        (H)  V         V                T
2)A         D V        V         K      T
3)S                  PVA            N
4)A                          A  .R  T
5)      H  V                        T
6)TPGCVPCVREGNASRCWVAMTPTVATRDGKLPATQLRRHIDLLVGSATLCS


1)
2)        I              D
3)                       D
4)
5)        I
6)ALYVGDLCGSVFLVGQLFTFSPRRHWTTQGCNCSI
```

SUMMARY: "S" AA117-308 (93%)

HCT#18, PBL-Th, EC1(Italy) have 97% homology with HCV-1
JH23 and JH 27 have 96% and 95% homology with HCV-1, respectively

<u>AA#300-438 (C-terminal region of the putative envelope region and amino ~1/3 of NS1)</u>

1) JH23   ?
2) JH27   ?
3) Japanese isolate (T. Miyamura)   ?
4) EC10 (Italy)   2 clones sequenced
                                   (one nt difference, which did not
                                   result in an amino acid change)
5) HCV-1 (chimpanzee)   multiple

                                   S ←——→ NS 1
1) D   A   V
2) D   A
3)   V S   VM   V
4)
5)TTQGCNCSIYPGHITGHRMAWDMMMNWSPTTALVMAQLLRIPQAILDMIAGA

1)   M   R   A RST A   VA
2)   T   YT   N   A R   TQALT   F
3)   L   Y   I M   GH R   VQ   VT   TLT
4)   A   I   AK   TASLTA
5)HWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDAETHVTGGSAGHTVSGFVSL

1)FS   R   I   I   T   V
2)FT   D I   I   R   A   D
3)FR   S   K I   V   I   R   Q   F
4)FNL   I   I   R   N
5)LAPGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWL

SUMMARY: NS 1 AA 330-660

"Isolate"	%Homology (AA330-438)	%Homology (AA383-405)
JH23	83	57
JH27	80	39
Japanese	73	48
EC10 (Italy)	84	48

FIG. 10